

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:14:56 ; Search time 113 Seconds  
(without alignments)  
5044.656 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 3820

Sequence: 1 atgcaaatgaattataaag.....aagagctgaagtgtatctaa 2139

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_n2p\_model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US1008355/runat\_16052003\_110337\_9308/app\_query.fasta\_1.2311  
-DB=A\_Geneseq\_101002 -QFMT=fastan -SUFFIX=n2p\_rag -MINMATCH=0 -1-LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US1008355\_CG\_N\_1153 @runat\_16052003\_110337\_9308 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_101002:\*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3719	97.4	712	23	AAO15205	Porphyromonas ging
2	274	7.2	52	23	AAO15206	Porphyromonas ging
3	174	4.6	1017	22	AAB59813	TutD protein #4.
4	174	4.6	1615	22	AAB59826	Protein #3 encoded
5	154.5	3.9	1592	22	AAB59827	Protein #4 encoded
6	153.5	3.9	999	22	AAB59817	TutD protein #8
7	153.5	4.0	1296	23	ABG66702	Human novel polype
8	153.5	4.0	1296	23	ABG66756	Human novel polype
9	148.5	3.9	1528	18	AAW20056	C. elegans UNC-53
10	148.5	3.9	1583	18	AAW20057	C. elegans UNC-53
11	148	3.9	2570	22	ABG06375	Novel human diagno
12	146.5	3.8	1368	23	ABP27518	Streptococcus poly
13	141.5	3.7	1795	22	ABB69806	Drosophila melanog
14	138	3.5	430	20	ABY31745	Mycobacterium tube
15	138	3.5	684	22	ABB61705	Drosophila melanog
16	135	3.5	1605	22	AAB59824	Protein #1 encoded
17	134.5	3.5	1022	22	AAB59814	TutD protein #5.
18	133	3.5	883	22	AAU37789	Streptococcus pneu
19	132	3.3	571	20	AAI36845	Protein involved i
20	130	3.4	883	20	AAI08339	S. pneumoniae val
21	130	3.4	883	21	AAI90514	Streptococcus pneu
22	130	3.4	883	23	AAU97883	Streptococcus pneu
23	130	3.4	1061	20	AAW87504	Human N-methyl-D-a
24	127.5	3.3	1194	18	AAW21725	Modified streptoki
25	127	3.2	388	20	AAI04998	Mycobacterium spec
26	127	3.3	883	22	AAU38091	Streptococcus pneu
27	127	3.3	1592	22	AAB59827	Protein #4 encoded
28	126.5	3.3	1370	23	ABP27517	Streptococcus poly
29	126.5	3.3	2870	21	AAI95559	Caenorhabditis ele
30	126.5	3.3	3178	21	AAI95556	Caenorhabditis ele
31	124	3.2	571	22	AAI27242	Human EXMAD-20 SEQ
32	123.5	3.2	713	20	AAI43380	M. catarrhalis str
33	122.5	3.2	26	23	AAO15221	Porphyromonas ging
34	122.5	3.2	711	17	AAI88649	Neisseria meningit
35	122.5	3.2	800	18	AAW21723	Plasminogen-bindin
36	122.5	3.2	813	18	AAW21728	Wild type plasmino
37	122.5	3.2	1181	18	AAW21727	Streptokinase/malt
38	122.5	3.2	1194	18	AAW21724	Modified streptoki
39	122.5	3.2	1194	18	AAW21726	Streptokinase/malt
40	122	3.1	677	22	AAU57954	Propionibacterium
41	121.5	3.2	746	22	ABB59201	Drosophila melanog
42	121	3.2	2432	21	AAI85565	Human homologue of
43	120	3.1	999	22	AAB59817	TutD protein #8
44	120	3.0	2467	23	ABB91303	Herbicidally activ
45	119.5	3.1	2042	19	AAW56319	Haemophilus paraga

ALIGNMENTS

RESULT 1  
AAO15205  
ID AAO15205 standard; Protein; 712 AA.  
XX  
AC AAO15205;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).  
XX  
KW Dipeptidylpeptidase-7; DPP-7; enzyme: amidolytic cleavage;  
DPP-7 inhibitor identification; periodontal disease; gingivitis;  
periodontitis.  
XX  
OS Porphyromonas gingivalis.  
XX  
PN WO200238742-A2.  
XX  
PD 16-MAY-2002.

XX 08-NOV-2001; 2001WO-US46782.  
XX 08-NOV-2000; 2000US-246827P.  
XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
PI Travis J, Potempa JS, Banbula A, Bugno M;  
XX N-PSDB; AAL43635.  
XX  
XX Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
PT the dipeptidylpeptidase for protecting an animal from periodontal  
PT disease caused by Porphyromonas gingivalis  
XX  
XX Claim 7; Fig 4; 65pp; English.  
XX  
XX The invention comprises the amino acid and coding sequence of the  
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
CC second and third amino acids from the N-terminal end of a target peptide.  
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a  
CC substituent on the alpha-carbon atom of the second amino acid from the  
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
CC reducing the growth of a bacterium and protecting an animal from a  
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
CC or periodontitis). The present amino acid sequence represents the  
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the  
CC invention.  
XX  
SQ Sequence 712 AA;

## Alignment Scores:

Pred. No.: 0 Length: 712  
Score: 3719.00 Matches: 712  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.36% Indels: 0  
DB: 23 Gaps: 0

US-10-008-355-1 (1-2139) x AA015205 (1-712)

QY	1	ATGCAATGAAATTAATAAGTATCTCTCGGAGCAGCCCTGCTGTGGGTGCTTCAGG	60
DB	1	MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuGlyAlaSerGly	20
QY	61	GTAGCCAAAGCCGACAAAGCATGTGGCTCCTCAACGAATCAATCAGGAGATCTGGAT	120
DB	21	ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnLeuAsnLeuAsp	40
QY	121	CGAATCGGTGAGCTCGGCTTTACGCTCCGCTGGGATTCGCTCTACAGTTTCGACAGCCG	180
DB	41	ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro	60
QY	181	TCCATTGCCAATGCCGTGGTATTCTTCGGTGGCGGATGTACGGTATACAGTGTCCGAT	240
DB	61	SerIleAlaAsnAlaValValIlePheGlyGlyCysThrGlyIleThrValSerAsp	80
QY	241	CAGGGCCTGATCTTTACCAACCACTACCGGATACGGTCTATCCAGAGCCCAAGCAGC	300
DB	81	GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr	100
QY	301	GTGGATCAGCACTATCTGCGGATGTTTCGTTCTCGCAGCATGGTGAGGCTTCGCG	360
DB	101	ValAspHisAspTyrLeuArgaspGlyPheValSerArgThrMetGlyGluLeuPro	120
QY	361	ATTCCGGGTCTTTCCGTGAAGTATCTGCCAAGATCGTGAAGGTACCGACAGGTAGAA	420
DB	121	IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu	140
QY	421	GGACAGCTCAAGGTATCACTGACGAGATGGAGCGTCTCGCAAGAGCTCAGGAGGTATGC	480

DB	141	GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys	160
QY	481	CRAGAACTGGCCCAAAAAGAAATGCAGACGAGAGCAACCACTCTGCATCTAGAGCCTTTC	540
DB	161	GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe	180
QY	541	TATTCCAAACAACGAATACTTCTCATCTGCTACGATGTATTCAAGGACGTTCTGATGGTA	600
DB	181	TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal	200
QY	601	TTTGCTCTCCAGCTCTGTAGTAAGTTCCGGAGCGGATACGAGCAACTGGATGTGGCCG	660
DB	201	PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTyrPro	220
QY	661	CGTCACACGGGCGACTTCAGCGTATTCCCGCTGTATCCCGGTGCCGACACCGCCGCC	720
DB	221	ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla	240
QY	721	GAATACAGCAAGGACAAATAAACCTATAAGCCGTTTACTTTCGCTGCCGTATCCATGCAA	780
DB	241	GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln	260
QY	781	GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGCAGTACGGATCCCTAC	840
DB	261	GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr	280
QY	841	CYCACHTCTGGGCTGTGAAGATCGTATCGAANAACGAGAACAACTCTCGTATCGAAGTT	900
DB	281	LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluVal	300
QY	901	CGCGGTATCAACGACGACATCTGGAAGAGCCATCAGCGCAGATCAGGCTACCGGTATC	960
DB	301	ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle	320
QY	961	AAATATGCCAGCAAGTATGCTCAGAGTGTAACTATTGGAAGAATTCGATCGGTATGAAC	1020
DB	321	LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn	340
QY	1021	CGCGGTCTCGCTGCTTTCAGCTGATAGTCTGAACGTCGCGGAGGAGGAGGATTCGCA	1080
DB	341	ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArgAlaPheAla	360
QY	1081	GACTGATCCGTAAAGAACGCAAGAGTGTCTATGGGCGATGATTGTCTCTCGAA	1140
DB	361	AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerLeuGlu	380
QY	1141	AAGGCTTATAAGGAAGGAGCCAAAGCCGAGATGACTTATTTGAGCGGAGCGCTC	1200
DB	381	LysAlaTyrLysGluGlyAlaLysAlaAsnArgLysMetThrTyrLeuSerGluThrLeu	400
QY	1201	TTCCGTTGTTACCGAGGTGGTTCGTTTTCGACAGTTTCCCAACGATTCGCTACAAATCCT	1260
DB	401	PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro	420
QY	1261	GATGCTCATCGCGGTATCTCAAAATCGCTTGACGACAAGTACAAAGCTACCTCCCTCG	1320
DB	421	AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer	440
QY	1321	CTCGACCGTAAAGTGTGTCGCCGCTATGTCGATATTTGTACGCCGCGGTATCCCTGCCGAC	1380
DB	441	LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp	460
QY	1381	AAGCTCCCGATATATTCAAGATGTAATCGACAAGAAATTCAAAGCCGACGAGAGAG	1440
DB	461	LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys	480
QY	1441	TATGACAGCTTCTGTTTCGACAGAGTGTGGTTCCTTATAGGACAAAGTTCCATGCCATG	1500
DB	481	TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet	500
QY	1501	CTCAAGCTCATGACAAAGGAAAGTTTTCGCAAGGCTATCGAAGAGATCCGCGCAGTAG	1560



Db 501 LeuLysSerMetAspLysGluLysPheAlaLysAlaLysAlaLysGluLysAspProAlaValGlu 520  
 QY 1561 CTTTCCAGAGCGTAATAGTGTCTGCGCGTATTTCAGGCGCGATGGCCAAATGCC 1620  
 Db 521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540  
 QY 1621 TATGCCATTGAGAGGCGAAGCGTCTTTCTTCCCGGTTTGGCTGAGATGATCCCGGGA 1680  
 Db 541 TyrAlaIleGluLysGlyLysArgLysPheAlaGlyLeuArgGluMetTyrProGly 560  
 QY 1681 CGTCTCTGCGGAGCGATGCAACTTCACCATGCTATGAGTACGGCTCCATCAAGGGA 1740  
 Db 561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580  
 QY 1741 TATGAACCGCAGGAGCGTGGCTGTACAACTATCATACGACAGCAAGGCGTATTGGAG 1800  
 Db 581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600  
 QY 1801 AAGCAGGATCCTAAGAGCGATGAGTTTCCCGTACAGGAGAATATCCTCGACCTCTTCGC 1860  
 Db 601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620  
 QY 1861 ACCAAAACATATGCTCGCTATGCGGAGAACGCTCAGCTCCATATCGCTTCTTATCGAAC 1920  
 Db 621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640  
 QY 1921 AACGACATACGCGGCGGTAACTCCGCTAGCCCGTATTTCGATAAGACGCGCGTCTGATC 1980  
 Db 641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660  
 QY 1981 GGTCTTCTTCGATGCGCAACTGGAGAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT 2040  
 Db 661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680  
 QY 2041 CTGACGCGCACAATCAGCGTGGACATCCGCTACGCTTCTTTCATGATTGACAAATGGGT 2100  
 Db 681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700  
 QY 2101 CAGTCCCCCGCTCTCATCCAAAGAGCTGAAGTTGATC 2136  
 Db 701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712

RESULT 2  
 AA015206  
 ID AA015206 standard; Protein; 52 AA.  
 XX AC AA015206;  
 XX DT 05-SEP-2002 (first entry)  
 XX DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.  
 XX KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
 XX KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
 XX KW periodontitis.  
 XX OS Porphyromonas gingivalis.  
 XX OS WO200238742-A2.  
 XX PN 16-MAY-2002.  
 XX PF 08-NOV-2001; 2001WO-US46782.  
 XX PR 08-NOV-2000; 2000US-246827P.  
 XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
 XX PI Travis J, Potempa JS, Banbula A, Bugno M;  
 XX DR WPI; 2002-490075/52.  
 XX PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of

the dipeptidylpeptidase for protecting an animal from periodontal disease caused by Porphyromonas gingivalis -  
 Example 6; Fig 5; 65pp; English.

The invention comprises the amino acid and coding sequence of the Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7 enzyme has amidolytic activity for cleavage of a peptide bond between the second and third amino acids from the N-terminal end of a target peptide. The DPP-7 target peptide has an aliphatic or aromatic residue as a substituent on the alpha-carbon atom of the second amino acid from the N-terminal end. The DPP-1 protein and DNA sequences of the invention are useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for reducing the growth of a bacterium and protecting an animal from a periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis or periodontitis). The present amino acid sequence represents the C-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme.

SQ Sequence 52 AA;

Alignment Scores:

Pred. No.: 8,98e-18 Length: 52  
 Score: 274.00 Matches: 52  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 7.17% Indels: 0  
 DB: 23 Gaps: 0

US-10-008-355-1 (1-2139) x AA015206 (1-52)

QY 1930 ACGGCGGTAACTCCGCTAGCCCGTATTTCGATAAGACGCGCTCTGATCGGCTTGTCT 1989  
 Db 1 ThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla 20  
 QY 1990 TTCATGCGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGACCCGATCTGCAGCGC 2049  
 Db 21 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 40  
 QY 2050 ACAATCAGCGTGGACATCCGCTACGCTTCTCTTCATG 2085  
 Db 41 ThrIleSerValAspIleArgTyrValLeuPheMet 52

RESULT 3

AA059813  
 ID AAB59813 standard; Protein; 1017 AA.

XX AC AAB59813;  
 XX DT 04-APR-2001 (first entry)  
 XX DE TutD protein #4.  
 XX KW Toluene degradation; enzyme; waste degradation; TutD.  
 XX OS Thauera aromatica.  
 XX OS Xanthomonas maltophilia.  
 XX OS Geobacter metallireducens.  
 XX OS Azococcus toluliticus.  
 XX PN WO200072650-A2.  
 XX PF 07-DEC-2000.  
 XX PR 24-MAY-2000; 2000WO-US14298.  
 XX PA 01-JUN-1999; 99US-0323872.  
 XX PI (UYOH-) UNIV OHIO.  
 XX DR Coschigano PW;  
 XX DR WPI; 2001-041080/05.

DR N-PSDB; AAF23625, AAF23627.  
 XX Composition comprising toluene degrading enzyme useful for biological  
 PT treatment of organic compounds, especially for degrading toluene or its  
 PT analogs -  
 XX  
 XX Disclosure; Fig 5; 122pp; English.  
 XX  
 CC The present invention relates to toluene degrading enzyme genes and  
 CC proteins tuth (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),  
 CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The  
 CC toluene degrading enzymes are homologues of pyruvate formate lyase. The  
 CC toluene degrading enzymes are useful for biological treatment of organic  
 CC compounds and in particular for the degradation of toluene and its  
 CC analogs contained in liquid or solid waste source. The present sequence  
 CC is a protein sequence for toluene degrading enzyme, Tutd.  
 XX  
 SQ Sequence 1017 AA;  
 Alignment Scores:  
 Pred. No.: 3.52e-07 Length: 1017  
 Score: 174.00 Matches: 176  
 Percent Similarity: 34.00% Conservative: 78  
 Best Local Similarity: 23.56% Mismatches: 275  
 Query Match: 4.55% Indels: 218  
 DB: 22 Gaps: 40

US-10-008-355-1 (1-2139) x AAB59813 (1-1017)

Qy 65 CCAAGCGCAAGGATGGTCTCAACGAACCTCAAGAGAGATCGATCGAA 124  
 Db 111 ProLysProArgProThrCysArgProSerProGlyThrAlaArgValSerThrThr 130  
 Qy 125 TCGGTGACCTGGCTTTACGCTCCGCTGG----- 154  
 Db 131 SerProArgSerThrGlyArgArgTrpSerSerProAlaArgArgSerAlaGlyArg 150  
 Qy 155 -----ATTGCGTCTACAGTTTCGACAAAGCGGCCATTG-----CCA 190  
 Db 151 AlaGlyArgAlaGlyCysAlaArgSerSerArgLysThrSerArgProIleArgSerAla 170  
 Qy 191 ATGCGGTGGTATCT-----TCGTGGGGATGTACCG----- 223  
 Db 171 ArgProSerCysSerLysSerProThrSerValSerAlaPheProProSerProAlaArg 190  
 Qy 224 GTATCACAGTGTCCGATCAGGCGCTGTATTACCAACCCACTGCGGATACGGTGCTA 283  
 Db 191 AlaSerArgThrArgCysArgArgAsnSerLeuProSerSerValThrArgSerSerAla 210  
 Qy 284 TCCAGAGCCAAAGCAGCGGTGATCACGACTATCTCGCGGATGGTTTCTCGCAGCA 343  
 Db 211 ThrArgAlaAlaThrProArgArgLysThrProCys----- 222  
 Qy 344 TGGGTGAGGAGTCCGATTCGGGTCTTCCGTGAAGTATCTCGCAAGATCGTGAAGG 403  
 Db 223 -----CysGlyArgThrArg 228  
 Qy 404 TAACGGACAAGGTAGAGGACAGCTCAAGGTATCACTGACGAGATGGAGGCTCGCGCA 463  
 Db 229 ProProSerSerThrArgAsnSerSerArgAlaThrTrpMetArgTrpAsnSerSerArg 248  
 Qy 464 AAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAAGAAATGCAGAGCAAGAACCACTCT 523  
 Db 249 TrpAsnValArgPheProSerMetAlaProAlaSerArgAlaProThrAlaLysSerSer 268  
 Qy 524 GCATCGTAGAGCCCTTATTCCCAACAGAACTACTCTCTCATCGTACCATGATTCA 583  
 Db 269 Arg-----GlyArgThrIleCysSerSerProSerProSerAlaAlaProThrProArgAla 286  
 Qy 584 AGGAGGTTTCGTATGTTCTCTCCAGCTCTGTAGGTAAAGTTCGAGGCGATACGG 643  
 Db 287 ArgThr-----ProAlaThrThrProThrProSerSerArgGln 299

Qy 644 ACAACTGGATGTGGCCGC-----GTCAACGCGGCCACT--- 676  
 Db 300 ProSerGlySerAlaArgProSerProProSerSerSerAlaIleProArgThrAla 319  
 Qy 677 -----TCAGGGTATTCGGGTGTGTATGCCGGTGGCG 706  
 Db 320 ArgArgArgCysAlaGlyPheSerSerAlaSerAlaThrAspSerAlaIleArgArgSer 339  
 Qy 707 ACAACCGCGCGCCCAATACAGCAAGGACAAATACCCCTATATAGCCCGCTTACTTCGCTG 766  
 Db 340 SerThrThrArgSerAlaArgSerArgAsnThrProSerSerAlaSerThrAlaThr 359  
 Qy 767 CCGTATCCATCAAG-----GCTACAAGCGCTGACGACTATGCCATGACCA 811  
 Db 360 AlaProProThrArgLysProThrThrLysSerThrCysAlaCysArgProAlaSer 379  
 Qy 812 TCGGTTTCC-----CGGGCAGTACGGATCGCTACCTACCTCTTCTT 850  
 Db 380 ThrValAlaAlaArgArgLysLysProValArgLysValAlaAlaGlnSerSerArgPro 399  
 Qy 851 GGGGTGTGGAAGATCGTATCGAAACAGAGAAACATCTCTCGTATCGAAGTTCGCGGTATCA 910  
 Db 400 SerCysTrpLysSerArgSerMetThrAlaThr----- 410  
 Qy 911 AGCAAGGATCTGGAAGGAGCCATGAGCG---CAGATCAGGCTACCCGTA-----TCA 961  
 Db 411 -----ThrGlyArgThrProThrCysAsnSerAlaArgArgProValIleSerArg 427  
 Qy 962 AATATGCCAGAGTATCTCAGAGTGTCTAATATTGGAAGAAATTCGATCGGTATGAACC 1021  
 Db 428 ArgSerProSerArgMet-----PheGlyArgLeu-----Ser 438  
 Qy 1022 CGGTCTCGCTCGCTTTCACGTGATAGTCTGATGCGGTGCGGAGGAAAGAGCATTCGCAG 1081  
 Db 439 AlaSerSerIleAsnMetArgSerThrSerValSerAlaProArgThrCysArgAla--- 457  
 Qy 1082 ACTGATCCGTGAAGACGCGAAGAGTGTCTATGCGGATGATTTGTCTCTCTCGAAA 1141  
 Db 458 ThrSerSerAlaSerCysArgLysLeuSer-----CysProGlnSerThr 473  
 Qy 1142 AGGCTTATAAGGAGGAG-----CCAAGGCCA----- 1168  
 Db 474 ThrAlaAlaTrpAsnSerGlyTrpThrProAlaProCysProSerSerProMetAlaGly 493  
 Qy 1169 ACCGTGAGATGACTATTGTTGAGCGAGAGCTTCG---GTGTACCGAGGTGGTTCGT 1225  
 Db 494 ThrThrArg-----SerArgSerSerArgArgThrProSerTrp----- 507  
 Qy 1226 TTGCACAGTTTGCACAG---CATTTGGCTACAAATCTCTGATGCTCATCCCGGTATCTCA 1282  
 Db 508 -----ProSerArgAsnTrp-----TyrSerArg 515  
 Qy 1283 AATCGCTTGACGACAGGTACAAAGACTACCTCCCTCGCTCGACCGTAAAGTGTGCGCG 1342  
 Db 516 ArgArgAsnThrProSerSerAsnSerAlaLysArgArgThrGlyLysValSerArg 535  
 Qy 1343 CCATCTCGATATTGTACGCGCGC-----GTATCCCTCGCG 1378  
 Db 536 LysCysAlaSerThrSerSerGlyArgArgSerGlyAlaThrThrMetIleThrProThr 555  
 Qy 1379 ACAAGTCCCGCATATATTCGAAGAAATGTAATCGACAAGAAATCAAGGCGACACGA--- 1435  
 Db 556 ValSerSerProAla-----SerThrArgLysSerSerAlaAlaLysCys 570  
 Qy 1436 ---AGAAGTATCGACATTCGTATTCGACAAGAGTGTCTCTTATAGCCACAGTTC 1492  
 Db 571 AlaArgSer---ProThrThrLeuValValArgSerCysArgLeuValArgLeuSerAla 589  
 Qy 1493 ATGCCATCTCAAGTCCATGCAAGGAAAGATTTGCCAAGG-----CTATCGAGA 1543  
 Db 590 -----CysThrTrpLysSerValArgAlaArgAlaProArgArgThrGlyAlaSerGly 607  
 Qy 1544 AAGATCCGGCAGTAGAGCTTTCCCAAGAGCGTAATAGTGTCTGCTCGCGCTATTCAGCGCG 1603

```
Db 608 valLysArgGln-----ThrThrAlaAlaPheLeuPro 618
QY 1604 ATCGGATGGCAATGCCATATGAGAGGCAAGCGCTCTTTCTTTGGCGGTTTGC 1663
Db 619 ---ThrTrpGluProThrArgArgGlyArgArgCysCysAlaArgCysProArgCys 637
QY 1664 GTGAGA-----TGTACCCCGGACGTGCTCTGCCAGCGATGCCCAACTTCA 1708
Db 638 ArgArgThrArgAlaThrCysThrSerAlaCysArgCysArgSer----- 653
QY 1709 CCATGGCTAGCTACGGCTCCATCAAGGATATGACCCGACGGTGCCTGTGACA 1768
Db 654 ---Cys-----AlaProSerMetAlaSerLysSerGlyThrArgThrArgLeuGlyThr 670
QY 1769 ACTATC-----ATACGACAG 1783
Db 671 IleIleSerIleMetPheSerSerMetSerSerAlaArgMetLysCysAlaLeuArgSer 690
QY 1784 GCAAGGCGTATTGGAGACAGGATCTTAAGACGGATGAGTTTGGCG----- 1831
Db 691 AlaAsnPro-----ArgSerThrMetIleLeuSerCysAlaPheProAlaThrAlaLeu 708
QY 1832 -----TACAGGAGAATATCTCGACCTCTCCGACCAAAACTATGGTGGCTATG 1882
Db 709 GlySerThrPheArgProMetGlyArgThrProSerSerProValArgAsnArgIleSer 728
QY 1883 CCGAGAACGCTAGCTCCATATCGCTTTCCTATCGAACACGACATCACGG-----GCG 1936
Db 729 AlaHisProIleSerSerThrSerLysSerArgThrSerHisSerArgGlyAlaAla 748
QY 1937 GTAACCTCGGTAGCCCGGTAT 1957
Db 749 SerArgProProLeuProTyr 755

RESULT 4
AAB59826
ID AAB59826 standard; Protein; 1615 AA.
XX
AC AAB59826;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #3 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azorarcus toluyticus.
XX
PN W0200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI: 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs.
XX
PS Disclosure; Fig 12; 122pp; English.
XX
```

```
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1615 AA;
Alignment Scores:
Pred. No.: 4,87e-07 Length: 1615
Score: 174.00 Matches: 176
Percent Similarity: 34.00% Conservative: 78
Best Local Similarity: 23.56% Mismatches: 275
Query Match: 4.55% Indels: 218
DB: Gaps: 40
US-10-008-355-1 (1-2139) x AAB59826 (1-1615)
QY 65 CCAAGCGCAGCAAGGCGATGTGGCTCTCAACGAACCAATCAGSAGAATCTGGATCGAA 124
Db 709 ProLysProArgProThrCysArgProSerProGlyThrAlaArgValSerThrThr 728
QY 125 TGGCTGAGCTCGGCTTTACGCTCCCGTTGG----- 154
Db 729 SerProArgSerThrThrGlyArgArgTrpSerSerProAlaArgArgSerAlaGlyArg 748
QY 155 -----ATTGCGCTACAGTTTCGACAAAGCCGTCACATG-----CCA 190
Db 749 AlaGlyArgAlaGlyCysAlaArgSerSerArgLysThrSerArgProIleArgSerAla 768
QY 191 ATCGCTGGTTATCT-----TCGTCGGCGATGTACCG----- 223
Db 769 ArgProSerCysSerLysSerProThrSerValSerAlaPheProSerProAlaArg 788
QY 224 GTATCACAGTGTCCGATCAGGCGCTGATCTTTACCAACACCACCTGCGGATACGCTGCTA 283
Db 789 AlaSerArgThrArgCysArgArgAsnSerLeuProSerSerValThrArgSerSerAla 808
QY 284 TCCAGAGCCCAAGCAGCGTGGATCAGCACTATCTGCGCGGATGTTGTTTCTCGCAGA 343
Db 809 ThrArgAlaAlaThrProArgArgLysThrProCys----- 820
QY 344 TGGGTGAGGAGCTTCCGATTCGCGGTCTTCCGTGAAGTATCTCGCAAGATCGTGAAGG 403
Db 821 -----CysGlyArgThrThrArg 826
QY 404 TAACGGACAAAGTAAAGGACAGCTCAAGGGTATCACTGACGAGATGAGCGTCTGCGCA 463
Db 827 ProProSerSerThrArgAsnSerSerArgAlaThrTrpMetArgTrpAsnSerSerArg 846
QY 464 AAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAGAAATGCAAGAGAGAACCAACTCT 523
Db 847 TrpAsnValArgPheProSerMetAlaProAlaSerArgAlaProThrAlaLysSerSer 866
QY 524 GCATCGTAGAGCCTTTCTATTCCCAACAGAACTACTTCCTCATCTCTACCATGATTCA 583
Db 867 Arg-----GlyArgThrIleCysSerSerProSerAlaAlaProThrProArgAla 884
QY 584 AGGACGTTCTGATGGTATTTGCTCTCCAGCTCTGTAGTAACTTCGAGCGCATACGG 643
Db 885 ArgThr-----ProAlaThrThrProThrProSerSerArgGln 897
QY 644 ACAACTGGATGTGCGCG-----GTCACACGGCGACT--- 676
Db 898 ProSerGlySerAlaAlaArgProSerProProSerSerSerAlaIleProArgThrAla 917
QY 677 -----TCAGCGTATTCGCGGTGTATCCCGGTCGG 706
Db 918 ArgArgArgCysAlaGlyPheSerSerAlaSerAlaThrAspSerAlaIleArgSer 937
```

Qy	707	ACAACCGCGCGCGGCAATACACGAAGGACATAAACCCCTATAAGCCCGTTTACTTCGCTGCTG	766
Db	938	SerThrArgSerAlaArgSerArgAsnThrProSerSerAlaSerThraIaThr	957
Qy	767	CCGTATCCATGCAAG-----GCTACAAGGCTGACGACTATGCCATGACCA	811
Db	958	AlaProProThrArgLysProThrThrGlySerThrCysCysAlaCysArgProAlaSer	977
Qy	812	TCGGTTTCC-----CGGCGAGTACGGATCGCTACCTCACTTCCT	850
Db	978	ThrValalaalaArgLysLysProValArgLysValaAlaGlnSerSerArgPro	997
Qy	851	GGGGTGTGGAAGATCGTATPCGAAAACGAGAACAACTCGTATCGAAGTTCCGGGTATCA	910
Db	998	SerCysTrpLysSerArgSerMetThrAlaThr-----	1008
Qy	911	AGCAAGGCATCTGGAAGAACCATGACGG--CAGATCAGGCTACCGGTA-----TCA	961
Db	1009	-----ThrGlyArgThrProThrCysAsnSerAlaArgArgProValIleSerArg	1025
Qy	962	AATATGCCACGAAGTATCTAGAGTGTAACTATGCGAAGAAATTCATCGGTATGAACC	1021
Db	1026	ArgSerProSerArgMet-----PheGlyArgLeu-----Ser	1036
Qy	1022	GCGGTCTCGCTCGTCTTTGACGTAGTGTGTAAGCTGCGGAGGAAGAGCATTCGCAG	1081
Db	1037	AlaSerSerIleAsnMetArgSerThrSerValSerAlaProArgThrCysArgAla---	1055
Qy	1082	ACTGGATCCGTAAAGACGCGCAGAGTGTGTCTATGGCGGATGTATTGCTCTCTCGAAA	1141
Db	1056	ThrSerSerSerAlaSerCysArgCysLeuSer-----CysProGlnSerThr	1071
Qy	1142	AGGCTTATAGGAAGGAG-----CCAAGGCCA-----	1168
Db	1072	ThraIalaTrpAsnSerGlyTrpThrProAlaProCysProSerSerProMetAlaGly	1091
Qy	1169	ACCGTGAGATGACTATTATTTGACGAGACGCTCTTCG---GTGGTACCGAGGTGGTTCGTT	1225
Db	1092	ThrThrArg-----SerArgSerSerArgArgThrProSerTrp-----	1105
Qy	1226	TTGCACAGTTTGCCAAGC---CATTTGGCTACAAATCTGTATGCTCATGCCGGTATCCTCA	1282
Db	1106	-----ProSerArgAsnTrp-----	1113
Qy	1283	AATCGCTTGACAGCAATACAAGACTACCTCCCTCGCTCGACCGTAAGGTGCTGCCCG	1342
Db	1114	ArgArgAsnThrProSerSerAsnSerAlaLysArgArgThrGlyLysValSerArg	1133
Qy	1343	CCATGCTCGATATTTGACCGCGC-----GATCCCTGCCG	1378
Db	1134	LysCysAlaSerThrSerSerGlyArgArgSerGlyAlaThrThrMetIleThrProThr	1153
Qy	1379	ACAAGTCCCGCATATATTCAGATGTATTCGACAGAAATTCGAAGGACACAGA---	1435
Db	1154	ValSerSerProAla-----SerThrArgLysSerSerSerAlaAlaLysCys	1168
Qy	1436	---AGAAGTATGCAGACTTCGTATTTCGACAGAGTGTGGTCTCTATATGCGCACAAGTTC	1492
Db	1169	AlaArgSer---ProThrThrLeuValValArgSerCysArgLeuValArgLeuSerAla	1187
Qy	1493	ATGCCATGCTCAAGTCCATGGACAGGAAAAGTTTGCCCAAGG-----CTATCCGAGA	1543
Db	1188	-----CysThrTrpLysSerValArgAlaArgAlaProArgArgThrGlyAlaSerGly	1205
Qy	1544	AAGATCCGGCTAGAGCTTTCCAAGACGGTAAATAGCTGCTCGCGCTATTACGCGCG	1603
Db	1206	ValLysArgGln-----ThrThrAlaAlaPheLeuPro	1216
Qy	1604	ATGCGATGGCCATGCCTATGCATTGAGAAGGGCAACGGCTCTTTTCTTTCCCGGTTTC	1663
Db	1217	---ThrTrpGluProThrArgArgGlyArgArgArgCysAlaArgCysProArgCys	1235
Qy	1664	GTGAGA-----TGTACCCCGGACGTGCTCTGCCGAGCGGATGCCAACATCTCA	1708

Db	1236	ArgArgThrArgAlaThrCysThrSerAlaCysArgCysArgSer-----			----	1251
Qy	1709	CCATGGCTAGCTACGGCTCCATCAAGGATATGAACCGACGACGGTGCCTGGTACA				1768
Db	1252	--Cys-----AlaProSerMetAlaSerLysSerGlyThrArgThrArgLeuGlyThr				1268
Qy	1769	ACTATC-----			-----ATACGACAG	1783
Db	1269	IleIleSerIleMetPheSerSerMetSerSerAlaArgMetLysCysAlaLeuArgSer				1288
Qy	1784	GCRAAGCGGTATTGGAGACGAGGATCCTAACAGCGCATGATTGCCG-----			-----	1831
Db	1289	AlaAsnPro-----ArgSerThrMetIleLeuSerCysAlaPheProAlaThrAlaLeu				1306
Qy	1832	-----TACAGGAGATATCCTCGACCTCTTCGGACCAACAAACTATGTCGCTATG				1882
Db	1307	GlySerThrPheArgProMetGlyArgThrProSerSerProValArgAsnArgIleSer				1326
Qy	1883	CCGAGAACGGTCAGCTCCATATCGCTTTCTTCATCAACAGACATCACGG-----			-----GCG	1936
Db	1327	AlaHisProIleSerSerSerThrSerLysSerArgThrSerHisSerArgGlyAlaAla				1346
Qy	1937	GTAATCCCGTAGCCCGGTAT			1957	
Db	1347	SerArgProProLeuProTyr			1353	
RESULT 5						
AAB59827						
ID	AAB59827 standard; Protein; 1592 AA.					
XX	AAB59827;					
XX	04-APR-2001 (first entry)					
XX	Protein #4 encoded by TutD/E gene.					
DE	Toluene degradation; enzyme; waste degradation; TutE; TutD.					
KW	Thaueria aromatica.					
OS	Xanthomonas maltophilia.					
OS	Geobacter metallireducens.					
OS	Azarcus toluylticus.					
XX	W0200072650-A2.					
XX	07-DEC-2000.					
XX	24-MAY-2000; 2000WO-US14298.					
XX	01-JUN-1999; 99US-0323872.					
XX	(UYOH-) UNIV OHIO.					
PA	Coschigano PW;					
PI	WPI: 2001-041080/05.					
XX	N-PSDB; AAF23627.					
DR	Composition comprising toluene degrading enzyme useful for biological					
PT	treatment of organic compounds, especially for degrading toluene or its					
PT	analog -					
XX	Disclosure; Fig 12; 122pp; English.					
FS	The present invention relates to toluene degrading enzyme genes and					
CC	proteins tutC (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832)					
CC	tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The					
CC	toluene degrading enzymes are homologues of pyruvate formate lyase. The					
CC	toluene degrading enzymes are useful for biological treatment of organic					
CC	compounds and in particular for the degradation of toluene and its					
CC	analog contained in liquid or solid waste source. The present sequenc					
CC	is a protein sequence encoded by toluene degrading enzyme gene, TutD/E					

XX SQ Sequence 1592 AA;  
 Alignment Scores: 3.74e-05 Length: 1592  
 Pred. No.: 154.50 Matches: 140  
 Score: 28.05% Conservativity: 46  
 Best Local Similarity: 21.12% Mismatches: 172  
 Query Match: 3.90% Indels: 305  
 DB: 22 Gaps: 38

US-10-008-355-1 (1-2139) x AAB59827 (1-1592)

QY 2080 AGAAGACTAGCGGATGCCAGGTGATGTGCGCTGCAGATCGGTTTCGAACCTCGATGT 2021  
 DB 768 ArgCysArgAspAlaCysGluArg  
 QY 2020 CACCACATCATGCTCCAGTTCGATCGCAAGCAAGACCGATCAGACGGCGGCTTCATAT 1961  
 DB 780 ProGlyProArgSerAlaProSerIleArgArgGlySerArgAspArgSerArgAlaSer 799  
 QY 1960 CGAATACGGGCTACCGGAGTTACGCCCGCTGATGT  
 DB 800 ArgSerArgGly  
 QY 1924 -----CCTGTTCTCATAGGA-----AAGCGA 1904  
 DB 816 ArgArgArgArgCysSerIleGlyAlaSerSerGlyCysProHisProProValArgArg 835  
 QY 1903 TATGGAGCTGACCGTTCGCGCATAGCAGCACCATTAGTTTTCGTCGGAGAGGTCGAGGA 1844  
 DB 836 SerProValAsnSerSerLysArgAlaHisArgArgCysThrAlaArgArgGlyArgPhe 855  
 QY 1843 TATTCTCCTGTA---CGGCAAACTCATCGCTTAGGATCCTGCTCTCCAATACGCCCT 1787  
 DB 856 ArgGlyProThrSerArgAspThrGlyArg  
 QY 1786 TGC-----CTG 1781  
 DB 868 CysTrpArgTrpProArgProArgCysArgCysSerArgArgTrpGlyArgProLeu 887  
 QY 1780 TCGTATGATAGTTGTACAGCAGCGCTCCTCGGTTTCATATCCCTGATGAGCGGTAGC 1721  
 DB 888 TrpAlaSerGlyCysProArgAlaArg  
 QY 1720 TCATACCGATCGTGAAGTTGGCATCGCTCGGCA  
 DB 901 Ser---AsnTrpSerSerGlyArgSerSerAlaAlaSerProLysArgThrCysGlyArg 919  
 QY 1684 CAGTCCGGGTACATCTCAGCAACCGGCAAGAAAGACGCTTGCCTTCTCAATGG 1625  
 DB 920 ArgValArgSerAspThrSerAlaArgArgSerArg  
 QY 1624 CATAGGCATTGCCATCCATCGCTGCAATAGCGGAGCAGCAGCTATTACGCTCTGG 1565  
 DB 934  
 QY 1564 AAAGCTCTACTCGCGGATCTTTCTCGATAGCTTGCGAAACTTTCTGTTCCTGACT 1505  
 DB 935  
 QY 1504 TGAGCATGGCATGAACCTGT---CGCTATAGGAACACACTCTTGTGCAATACGAAGT 1448  
 DB 942  
 QY 1447 CTGCATCTCTCTGTCGCGCTTTGATTTCTTGTTCGATTACATTCTTGTGATATATCGG 1398  
 DB 958 ThrCysThrAlaArgCysGly  
 QY 1387 GGAGCTTGTGCGGAGGATACCGCGGCTACAAATATCGAGCATGCGCG------GCA 1337  
 DB 966 AspGlyCysSerAlaPhePheGlyAsnProLeuHisArgSerLeuArgGlyProTrpAla 985

QY 1336 GCACCTTACGCTGAGCGAGGAGGAGGTAGTCTTTGTACTTCTGCTCAAGCGATTGAGGA 1277  
 DB 986 AlaProPhe---ArgAlaHisArgSerArgSerThrThrArgCysAlaValArgGly 1004  
 QY 1276 ---TACCGCATGAGCAT-----1262  
 DB 1005 SerSerArgHisAspArgThrAlaSerThrArgArgProHisLysProProLysGlyCys 1024  
 QY 1261 ---CAGGATTTGTAGCCAAATCGCTGG-----CAAATGTCGCAAAAGCAA-----1220  
 DB 1025 AlaThrAspIleHisSerGlyArgTyrcysTrpProArgThrAlaSerSerArgAlaAla 1044  
 QY 1219 -----CCACCTCGGTACCAC 1205  
 DB 1045 SerGlyAlaSerAlaLysArgThrArgLeuArgArgArgSerCysProValArgSerPro 1064  
 QY 1204 CGAAGA-----GGTCTCGCTCAAAATAAGTCA 1178  
 DB 1065 ArgArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySerSerArgArgPro 1084  
 QY 1177 TCTCACGGTTGGCTTGGCTCTCTCTCTAT-----1148  
 DB 1085 SerSerGlyArgProTrpSerValProIleArgProSerSerIleCysGlyArgAlaVal 1104  
 QY 1147 -----AAGCCTTTTCGAGAGAAGACAATACATCGC 1118  
 DB 1105 GlyLeuThrSerProSerSerProLeuAsnArgProPheAlaArgArgSerAlaProAla 1124  
 QY 1117 CAT-----AGACAGCAGCTCTTGGCGTTCCTTACGATCCAGTCTCGCAATGCTCTTT 1067  
 DB 1125 SerThrProCysArgArgHisAsnArgArgTyrglySer-----ArgArgProPhe 1142  
 QY 1066 CCTCGGCAC-----GCTTACGACCTATCATCGTCAAGCAGCAGGA 1028  
 DB 1143 ArgArgArgPheAlaCysSerTrpSerSerGlnHisAspProAlaSerGlnAspProGln 1162  
 QY 1027 GACCGCGTTCATACCGATCGAATTCCTCAATAGTTAGCAGCTCTGAGCATACTTGTCTGG 968  
 DB 1163 ---ArgGly-----ThrCys---1166  
 QY 967 CATATTGATACGGGTAGCTGATCGCTCA-----TGCTTCTCTCCAGA 920  
 DB 1167 -----ProLeuArgAsnAlaCysProGlyTrpAlaPro-----1177  
 QY 919 TGCCTTGTCTGATACCGCAACTTCGATACGAGGATTTCTCTCGTTTTCGATACGATCTT 860  
 DB 1178 -----ArgAlaSerArg 1181  
 QY 859 CCACAC---CCCAAGAGTGTAGGATCCGCTACTGCCCG-----821  
 DB 1182 ProHisLeuProLeuArgArgCysLysGluArgCysProPheArgCysSerProAla 1201  
 QY 820 -----GGAAACCGATGTCATGGCATAGTCTGACGCTTGT-785  
 DB 1202 SerProProThrAlaSerProThrTrpProAlaSerSerGlySerProCysGlyAlaSer 1221  
 QY 784 -----AGCCTTTCATGATGATCGGACGCA-----761  
 DB 1222 GlyAlaSerThrProAlaSerAlaSerTrpAlaHisSerArgPheArgSerSerThrCys 1241  
 QY 760 -----ACTAAACGGCTTATAGGTTTAT-----TGTCCT-----731  
 DB 1242 SerSerLysAlaArgAspSerTyrglyAspHisGluThrHisArgCysProHisPheAla 1261  
 QY 730 -----TCCTGATTCGCGCGCGCTTGTGGCACCGCATACACGGGAATACGC 680  
 DB 1262 ArgGlnAlaAlaArgValArgSerAlaGlyHisAspArgAlaArgLysArgSerCysArg 1281  
 QY 679 TGAAGTCGCGCTGTGACCGCGCCACATCCAGTTCT-----CCGTAT 638  
 DB 1282 GluThrArgLeuCysGluSerArgHisAlaSerCysTrpArgGlyAspProValProAsn 1301  
 QY 637 CGCCTCCGA 629







Qy	1555	-----GTAGACCTTTCCAAGAGCGCTAATAGCTGCTGCTC	1588
Db	955	alGlySerAlaSerIleThrGlyPheProSerLeuSerSerSerAlaThrThrSerThrS	975
Qy	1589	GCCTATTACAGCGCGATGCCATGGCCAAATGCCTATGCATTGAGAAGGCGCAAGCTCTTT	1648
Db	975	erSerThrSerSerSerLeuThrThrAlaLeuThrGluIleThrPro-PheSerThrIle	994
Qy	1649	TCCTTTCGCGGTTTTCGCTGAGATGATACCCCGGACGGT-----CTCTGCCGAGCGATG	1699
Db	995	SerLeuProSerThrThrProCysProGlyThrIleThrIleThrIleValProAlaSer	1014
Qy	1700	CCAACTTCACCATCGGTA-----TGAGTACGGCTCCA-----	1732
Db	1015	ProThrAspProCysValGluMetAspProSerThrGluAlaThrSerProProThrThr	1034
Qy	1732	-----	1732
Db	1035	ProLeuThrValPheProPheThrThrGluMetValThrCysProThrSerIleSerIle	1054
Qy	1732	-----	1732
Db	1055	GlnThrThrLeuThrThrTyrMetAspThrSerSerMetMetProGluSerGluSerSer	1074
Qy	1733	-----TCAAGGATATGACACCCAGCAGCGTGCCTGGTACAACTATCATTA	1777
Db	1075	IleSerProAsnAlaSerSerSerThrGlyThrValProThrAsnThrValPhe	1094
Qy	1778	CGACAGGCAAGGCGGTATTGGAGACAGGATCCTAAGACGATGAGTTCCCGTACAGG	1837
Db	1095	ThrSerThrArg-----LeuPro-----	1100
Qy	1838	AGAAATPCTTCGACCTCTTCCGCACCAAAACTATGGTCGTATGCCGAGAAGC---	1894
Db	1101	--ThrSerGluThrTrpLeuSerAsnSerSerValIleProLeuProLeuProGlyVal	1119
Qy	1895	AGCTCCATATCGCTTTCCTATCGAACACGACATCACGGGCGGTAACTCCGGTACGCCCG	1954
Db	1120	SerThrIleProLeuThrMetLysProSerSer-----LeuPro	1133
Qy	1955	TATTGATAAGAACGGCGCTCTGATCGGTCCTTTCGATGGCAACTGGGAAGCTATGA	2014
Db	1134	ThrIleLeuArgThrSerSerLysSerThrHisProSerProThrThrArg-----	1151
Qy	2015	GTGCTGACATCGAGTTTCCGAACCCGATCTGCACGCCACAAATCAGCGTGGACATCCGCTACG	2074
Db	1152	-----ThrSerGluThrProValAlaThrThrGln-----ThrProThrThr	1165
Qy	2075	TTCTCTCATGATTGACAAATGGGTCAGTGCGCCCGCTCATCCAAGAGC	2125
Db	1166	LeuThrSerArgArgThrThrArgIleThrSerGlnMetThrThrGlnSer	1182
RESULT	8		
ABG66756			
ID	ABG66756	standard; Protein; 1296 AA.	
XX	ABG66756;		
XX			
DT	30-AUG-2002	(first entry)	
XX			
DE	Human novel polypeptide #91.		
XX			
KW	Human; inflammatory condition; shock; sepsis; immune response;		
KW	cancer; wound healing; central nervous system disease; haematopoiesis;		
KW	peripheral nervous system disease; amyotrophic lateral sclerosis; tend;		
KW	myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone		
KW	cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis		
KW	bone degenerative disorder; periodontal disease; reperfusion injury;		
KW	lung fibrosis; liver fibrosis; autolimmune disorder; bacterial infection;		
KW	allergic condition; thrombolysis; thrombosis; coagulation disorder;		
XX	fungal infection.		
XX			



Db 935 alPheSerThrThrIleHisSerValProSerSerProTyrIlePheSerThrGluAsnV 955  
 Qy 1555 -----GTAGAGCTTTCAAGAGCGTAATAGCTGCTGCTC 1588 .  
 Db 955 alGlySerAlaSerIleThrGlyPheProSerLeuSerSerAlaThrThrSerThrs 975  
 Qy 1589 GCGTATTACGCCGATCGGATGCCAATGCCTATGCCATTGAGAGGCGAAGGCTCTTT 1648  
 Db 975 erSerThrSerSerLeuThrThrAlaLeuThrGluIleThrPro-PheSerTyrIle 994  
 Qy 1649 TCTTTGCGCGTTCGGTCAGATGACCCCGGAGCTG-----CCTGCGGAGCCGATG 1699  
 Db 995 SerLeuProSerThrThrProCysProGlyThrIleThrIleThrIleValProAlaSer 1014  
 Qy 1700 CCAACTTCACCATCGGTA-----TCAGCTACGGCTCCA----- 1732  
 Db 1015 ProThrAspProCysValGluMetAspProSerThrGluAlaThrSerProProThrThr 1034  
 Qy 1732 ----- 1732  
 Db 1035 ProLeuThrValPheProPheThrThrGluMetValThrCysProThrSerIleSerIle 1054  
 Qy 1732 ----- 1732  
 Db 1055 GlnThrThrLeuThrThrTyrMetAspThrSerSerMetMetProGluSerGluSerSer 1074  
 Qy 1733 -----TCAAGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATATA 1777  
 Db 1075 IleSerProAsnAlaSerSerSerThrGlyThrGlyThrValProThrAsnThrValPhe 1094  
 Qy 1778 CGACAGGCAAGGCGGTATTGGAGAGCAGGATCCTAAGAGCGATCAGTTCCCGTACAGG 1837  
 Db 1095 ThrSerThrArg-----LeuPro----- 1100  
 Qy 1838 AGAATATCTCGACCTCTTCGCGACCAAAACTATGGTGTATGCCGAGAACG---GTC 1894  
 Db 1101 ---ThrSerGluThrThrTrpLeuSerAsnSerSerValIleProLeuProGlyVal 1119  
 Qy 1895 AGCTCCATATCGCTTTCCTATCGAACACGACATCACGGCGGTAACTCCGGTAGCCCGC 1954  
 Db 1120 SerThrIleProLeuThrMetLysProSerSer-----LeuPro 1133  
 Qy 1955 TATTCGATAAGAGCGCGTCTGATCGCTCTTGTCTTTCGATGGCAACTGGAAGCTATGA 2014  
 Db 1134 ThrIleLeuArgThrSerSerLysSerThrHisProSerProThrThrArg----- 1151  
 Qy 2015 GTGGTGACATCGAGTTCGAACCCGATCGCAGCGCACATCAGCGTGACATCCGCTACG 2074  
 Db 1152 -----ThrSerGluThrProValAlaThrThrGln-----ThrProThrThr 1165  
 Qy 2075 TTCTTTCATGATTGACAAATGGGTGATGCGCCCGCTCATCCAGAGC 2125  
 Db 1166 LeuThrSerArgThrThrArgIleThrSerGlnMetThrThrGlnSer 1182  
 RESULT 9  
 AAW20056  
 ID AAW20056 standard; protein; 1528 AA.  
 XX  
 AC AAW20056;  
 DX  
 DT 10-SEP-1997 (first entry)  
 XX  
 DE C. elegans UNC-53 protein variant 8A.  
 XX  
 KW UNC-53; neuronal regeneration; revascularisation; wound healing;  
 KW neurodegenerative disease; Alzheimer's disease; Huntington's;  
 KW peripheral neuropathies; metastasis inhibition; cancer.  
 XX  
 OS Caenorhabditis elegans.  
 XX  
 PN W09638555-A2.  
 XX  
 PD 05-DEC-1996.

XX 31-MAY-1996; 96WO-EP023111.  
 XX 31-MAY-1995; 95GB-0010944.  
 PR  
 XX (BOGA/) BOGAERT T.  
 PA (STRI/) STRINGHAM E.  
 PA (VAND/) VANDEKERCKHOVE J.  
 XX  
 PI Bogaert T, Stringham E, Vandekerckhove J;  
 XX WPI; 1997-034369/03.  
 DR N-PSDB: AAT71314.  
 XX  
 XX Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to  
 promote neuronal regeneration, revascularisation or wound healing.  
 PS Claim 20; Page 106-111; 278pp; English.  
 XX  
 CC UNC-53 protein variants 8A and 7A of Caenorhabditis elegans are new.  
 CC The UNC-53 proteins and nucleic acids are useful as medicaments to  
 CC promote neuronal regeneration, revascularisation or wound healing, or  
 CC for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or  
 CC Huntington's disease) or acute traumatic injuries. Transgenic cells and  
 CC organisms transfected with UNC-53 cDNA can be used to determine whether  
 CC a substance is an inhibitor or enhancer of the regulation of cell shape  
 CC or motility or the direction of cell migration by screening for a  
 CC phenotypic change in the cell. Inhibitors can be used to alleviate the  
 CC spread of disease inducing cells or metastasis. Probes derived from the  
 CC cDNA sequences can be used to identify homologues of the C. elegans  
 CC unc-53 gene. The UNC-53 protein can be used to identify proteins which  
 CC are active in the signal transduction pathway that can be used as  
 CC mentioned above.  
 XX  
 SQ Sequence 1528 AA;  
 Alignment Scores:  
 Pred. No.: 0.000139 Length: 1528  
 Score: 148.50 Matches: 106  
 Percent Similarity: 37.48% Conservative: 84  
 Best Local Similarity: 20.91% Mismatches: 211  
 Query Match: 3.89% Indels: 107  
 DB: 18 Gaps: 19  
 US-10-008-355-1 (1-2139) x AAW20056 (1-1528)  
 Qy 226 ATCAGAGTGTCCGATCAGGGGCTGATC-----TTTACCACCAACCACTGC 270  
 Db 1 MetThrThrSerAsnValGluIleProIleTyrThrAspTrpAlaAsnArgHisLeu 20  
 Qy 271 GGATACGGTGTATC--CAGAGCCAAAGCACGGTGGATCAGCTATCTCGCGATGGT 327  
 Db 21 SerLysGlySerLeuSerLysSerIleArgAspIleSerAsnAspPheArgAspTyrArg 40  
 Qy 328 TTCGTTTTCGACAGATGGGTGAGGAGCTTCCGATTCGGGTCTTTCGCGTAAGTATCTG 387  
 Db 41 LeuValSerGlnLeuIleAsnValIleValProIleAsnGluPheSerProAlaPheThr 60  
 Qy 388 CGCAAGATCGTCAAGTAACCGCAAGGTAGAACAGCTCAAGGGTATCCTCAGTACCAG 447  
 Db 61 LysArgLeuAlaLysIleThrSerAsnLeuAspGly---LeuGluThrCysLeuAspTyr 79  
 Qy 448 ATGGAGCGCTCTCGCAAGAGCTCAGGAGCTATGCCAAGACTGGCCAAAAGAAATGCA 507  
 Db 80 LeuLysAsnLeuGlyLeuAsp-----CysSerLysLeuThrLysThrAspIleAsp 96  
 Qy 508 GACGAGAACCAACTCTGCATCGTAGAGCCTTCTATTATCCAAACAAGTAATCTTCTCATC 567  
 Db 97 SerGlyAsnLeuGlyAlaValLeuGlnLeuLeuPheLeuLeuSerThrTyrLysGlnLys 116  
 Qy 568 GTCTACGATGTATTCAAGGAGCTTCGTATGGTATTTGCTCTCCAGCTCTGTAGTAGAG 627  
 Db 117 LeuArgGlnLeuLysLysAspGlnLysLeuGlnLeuProThrSerIle-MetPr 136

```
QY 628 TTCGAGCGGATACGGACAACTGGATGTGCGCGGTACACGGCGACTTCAGCGTATTC 687
Db      ::
136 oProAlaValSerLys----- 141
QY 688 CGCGTGATCGCGTCCGACACCGCGCGCCGCAATACAGCAGGACAAATAACCCCTAT 747
Db      ::::: ||| ||||| |||||
142 -----LeuProSerProArgValAlaThrSerAlaThrAlaSerAlaThrAsnProAs 159
QY 748 AAGCCCTTTACTTCGTCGCGGTATCCATCCAGGCTACAGGCTACAGGCTACGACTATGCCATG 807
Db      ||| ||| ::::: ||||| |||||
159 nSerAsnPheProGlnMet-----SerThrSerArgLeuGlnThr----- 172
QY 808 ACCATCGGTTCCGGCGCAGTACGGATCGTACTCTCTCTGGGTGTGGAAGATCGT 867
Db      ||||| ::::: ||||| :::::
173 ---ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyLeuProLysTh 191
QY 868 A-----TCGAAACGAGACAACTCTCGTATCGAAGTTCCG 903
Db      ||| ||| ::::: |||||
191 rSerGlyLeuLysProProSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211
QY 904 GGTATCAAGCAAGCATCTGGA----- 925
Db      ||||| ::::: |||||
211 gProSerSerArgSerSerGlyAsnAsnValGlySerThrIleSerThrSerAlaLy 231
QY 926 -----AGGAAGCCATGAGCGCA 942
Db      ||||| ::::: |||||
231 sSerLeuGluSerSerSerThrThrSerIleSerAsnLeuAsnArgProThrSerGl 251
QY 943 GATCAGGCTACCGGTATCAATATGCCAGCAAGTATCTCAGAGTCTAACTATTGGAAG 1002
Db      ::::: ||| ||| ::::: |||||
251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrLy 270
QY 1003 AATTCGATCG-----GTATGAACCGCGTCTCGCTCTGACGTATAGGT 1050
Db      ||| ||| ||||| |||||
270 sIleGlySerSerLysLeuAlaAlaProLysAlaValSerThrProLysLeuAlaSerVa 290
QY 1051 CGTAAGCGTCCGAGGAAGACATTCGACAGCTGGATCGGTAAAGACGGCAAGAGTCT 1110
Db      ::::: |||||
290 lLysThrIleGlyAlaLysGlnGluProAspAsnSerGlyGlyGlyGlyMetLe 310
QY 1111 GTCATGCGGATGATTGCTTCTCTCGAAAGGCTTATAGGAAGGAGCCAGCCCAAC 1170
Db      ::::: |||||
310 uLysLeuLeuLysPheSerSerLysAsnProSerSerSerSerAsnSerProGlnProTh 330
QY 1171 CGTGAGATGACTTATTGAGCGAGCAGCTCTTCGGTGGTACCGAGGTGGTTCGTTTGA 1230
Db      ::::: |||||
330 r-----ArgLysAlaAlaAlaValProGlnGlnThrLeuSe 343
QY 1231 CAGTTTCCCAACGATTGGTACAAATCCTGATGCTCATGCCGTATCCTCAAAATCGCTT 1290
Db      ::::: |||||
343 rLysIleAlaAla-----ProValLysSerGlyLeuLy 354
QY 1291 GACGACAGTACAAAGACTACCTCCCTCGCTCGACCGGTAGGTGCTCCCGCCATGCTC 1350
Db      ::::: |||||
354 sProProThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373
QY 1351 GATATTGTACGCGCGGTATCCCTCGCGACAAAGCTCCCGCATATATTCAAGATGTAATC 1410
Db      ::::: |||||
373 sValSerTyArg-----LysThrAspAlaProIleIleSerGlnGlnAspSe 389
QY 1411 GACAAGAAATTCAAAGGGACACGAAGAAG-----TATCGACACTTCGTATTTCGACAA 1463
Db      ||| ||| ::::: |||||
389 rLysArgCysSerLysSerSerGluGluSerGlyTyAlaGlyPhe-----As 406
QY 1464 GAGTGTGTTCTTATAGGACAAAGTCCATGCCATGCTCAAG---TCCATGGACAGGA 1520
Db      ||| ||| |||||
406 nSerThrSerProThrSerSerThrGluGlySerLeuSerMetHisSerThrSerSe 426
QY 1521 AAAGTTTCCCAAGGCTATCGAAGAAGATCCGGCA-----CTAGAGCTTTCCCAAGAG 1571
Db      ||| ::::: ||||| |||||
426 rLysSerSerThrSerAspGluLysSerProSerSerAspAspLeuThrLeuAsnAlaSe 446
```

```
QY 1572 CGTAATAGCTGCTGCTCGC 1590
Db      ::::: ||| |||
446 rIleValThrAlaIleArg 452

RESULT 10
AAW20057
ID AAW20057 standard; Protein; 1583 AA.
XX
AC AAW20057;
DT 10-SEP-1997 (first entry)
XX
DE C. elegans UNC-53 protein variant 7A.
XX
KW UNC-53; neuronal regeneration; revascularisation; wound healing;
neurodegenerative disease; Alzheimer's disease; Huntingdon's;
peripheral neuropathies; metastasis inhibition; cancer.
XX
OS Caenorhabditis elegans.
XX
PN WO9638555-A2.
XX
PD 05-DEC-1996.
XX
PF 31-MAY-1996; 96WO-EP02311.
XX
PR 31-MAY-1995; 95GB-0010944.
XX
PA (BOGA/) BOGAERT T.
(STRL/) STRINGHAM E.
(VAND/) VANDEKERCKHOVE J.
XX
PI Bogaert T, Stringham E, Vandekerckhove J;
DR WPI: 1997-034369/03.
DR N-PSDB; AAT71315.
XX
PT Caenorhabditis elegans UNC-53 protein 8A and 7A variants - useful to
promote neuronal regeneration, revascularisation or wound healing.
XX
PS Claim 22; Page 111-116; 278pp; English.
XX
CC UNC-53 protein variants 8A and 7A of Caenorhabditis elegans are new.
The UNC-53 proteins and nucleic acids are useful as medicaments to
promote neuronal regeneration, revascularisation or wound healing, or
for treatment of chronic neurodegenerative diseases (e.g. Alzheimer's or
Huntingdon's disease) or acute traumatic injuries. Transgenic cells and
organisms transfected with UNC-53 cDNA can be used to determine whether
a substance is an inhibitor or enhancer of the regulation of cell shape
or motility or the direction of cell migration by screening for a
phenotypic change in the cell. Inhibitors can be used to alleviate the
spread of disease inducing cells or metastasis. Probes derived from the
cDNA sequences can be used to identify homologues of the C. elegans
unc-53 gene. The UNC-53 protein can be used to identify proteins which
are active in the signal transduction pathway that can be used as
mentioned above.
XX
SQ Sequence 1583 AA;

Alignment Scores:
Pred. No.: 0.000142 Length: 1583
Score: 148.50 Matches: 106
Percent Similarity: 37.48% Conservative: 84
Best Local Similarity: 20.91% Mismatches: 211
Query Match: 3.89% Indels: 107
DB: 18 Gaps: 19

US-10-008-355-1 (1-2139) x AAW20057 (1-1583)
QY 226 ATCAGAGTGTCCGATCAGGGGCTGTATC-----TTTACCAACCACCTGC 270
Db      ::||| ||||| ||||| |||||
1 MethThrSerAsnValGluLeuIleProIleThrAspTrpAlaAsnArgHisLeu 20
```

```
Qy 271 GGATACGGTGTCTATC---CAGAGCAAGACGCGGTGGATCAGCAGTATCTCGCGGATGGT 327
Db 21 SerLysGlySerLeuSerLysSerIleAlaValLeuGlnLeuPheArgAspTyrArg 40
Qy 328 TTCGTTTCTGCAGCATGGGTGAGGAGCTTCCGATTCCTGGGTCTTTCCGTCGAAGTATCTG 387
Db 41 LeuValSerGlnLeuIleAsnValIleValProIleAsnGluPheSerProAlaPheThr 60
Qy 388 CGCAAGATCGTGAAGTAAAGCAAGGTAGAAGCAGCTCAAGGCTCAAGGCTATCACTGACGAG 447
Db 61 LysArgLeuAlaLysIleThrSerAsnLeuAspGly---LeuGluThrCysLeuAspTyr 79
Qy 448 ATGAGCGCTCTGCCAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAGAAAGAAATGCA 507
Db 80 LeuLysAsnLeuGlyLeuAsp-----CysSerLysLeuThrLysThrAspLeuAsp 96
Qy 508 GACGAGAACCAACTCTGCATCGTAGAGCCTTCTTATTCACACAGCAATACTTCTCTCATC 567
Db 97 SerGlyAsnLeuGlyAlaValLeuGlnLeuPheLeuLeuSerThrTyrLysGlnLys 116
Qy 568 GTCTACGATGTTATCAAGGAGCTTCGTATGTTATTTCTCTCCAGCTCTGTAGGTAAAG 627
Db 117 LeuArgGlnLeuLysLysAspGlnLysLysLeuGlnLeuProThrSerIle-MetPr 136
Qy 628 TTCGAGCGCATACGGACAACACTGGATGTGCGCGCTCACAGCGCGACTTCAGCGTATTC 687
Db 136 oProAlaValSerLys----- 141
Qy 688 CGCGTGTATGCCGTGCGGACACCGCGCGCGAATACAGACAGGACAATAAACCCCTAT 747
Db 142 -----LeuProSerProArgValAlaThrSerAlaThrAlaSerAlaThrAsnProAs 159
Qy 748 AAGCCGTTTACTTCGTCGCGTATCCATGCAAGCTACAGGCTGACGACTATGCCATG 807
Db 159 nSerAsnPheProGlnMet-----SerThrSerArgLeuGlnThr----- 172
Qy 808 ACCATCGGTTTCCCGGCAGTACGATCGTACCTACTTCTGGGTGTGGAAGATCGT 867
Db 173 ----ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyIleLysProLysTh 191
Qy 868 A-----TCGAAACGAGACAATCCTCGTATCGAAGTTGCG 903
Db 191 rSerGlyLeuLysProProSerSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211
Qy 904 GGTATCAAGCAAGCATCTGGA----- 925
Db 211 gProSerSerArgSerSerGlyAsnAsnValGlySerThrIleSerThrSerAlaLys 231
Qy 926 -----AGGAAGCCATGAGCGCA 942
Db 231 sSerLeuGluSerSerSerThrTyrSerSerIleSerAsnLeuAsnArgProThrSerGl 251
Qy 943 GATCAGGCTACCGTATCAAAATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAG 1002
Db 251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrLys 270
Qy 1003 AATTCGATCG-----GTATGAACCGCGTCTCGCTCTCTTACGCTGATAGGT 1050
Db 270 sIleGlySerSerLysLeuAlaAlaProLysAlaValSerThrProLysLeuAlaSerVa 290
Qy 1051 CGTAAGCGTCCGAGGAAGACATTCGACAGCTGGATCCGTAGACAGCGCAAGAGTGTCT 1110
Db 290 lLysThrIleGlyAlaLysGluProAspAsnSerGlyGlyGlyGlyGlyMetLe 310
Qy 1111 GTCATCGCATGATTCTCTCTCGAAAGGCTTATAGGAAGGAGCCAGCCCAAC 1170
Db 310 uLysLeuLysLeuPheSerSerLysAsnProSerSerSerSerAsnSerProGlnProTh 330
Qy 1171 CGTGAGATGACTTATTTGAGCGAGCAGCTCTTCGCTGTCACCGAGGTGGTTCGTTTTCGA 1230
Db 330 r-----ArgLysAlaAlaValProGlnGlnThrLeuSe 343
Qy 1231 CAGTTTCCCAACGATTTGGGTACAAAATCCTGATGCTGATGCGCGGTATCTCCTCAAAATCGCTT 1290
```

```
Db 343 rLysIleAlaAla-----ProValLysSerGlyLeuLy 354
Qy 1291 GACGACAAGTACAAGACTACCTCCCTCGCTGACCGCTAAGGTGCTGCCGCCATCCTC 1350
Db 354 sProThrThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373
Qy 1351 GATATTCTACGCGCGCGTATCCCTGCGCAGCAAGCTCCCGCATATATTCAAGAATGTATC 1410
Db 373 sValSerTyrArg-----LysThrAspAlaProIleIleSerGlnGlnAspSe 389
Qy 1411 GACAAGAAATTCAAAGCGCACAGGAAG-----TATCGAGACTTCGTATTCGACAA 1463
Db 389 rLysArgCysSerLysSerSerGluGluSerGlyTyrAlaGlyPhe-----As 406
Qy 1464 GAGTGTGTTCTTATAGCGACAGTTCCATGCCATCTCAAG---TCCATGGACAAGGA 1520
Db 406 nSerThrSerProThrSerSerSerThrGluGlySerLeuSerMetHisSerThrSerSe 426
Qy 1521 AAAGTTTGCACCAAGGCTATCGAGAAAGATCCGGCA-----GTAGAGCTTTCCAAAG 1571
Db 426 rLysSerThrSerAspGluLysSerProSerSerAspLeuThrLeuAsnAlaSe 446
Qy 1572 CGTAATAGCTGCTCCTCGC 1590
Db 446 rIleValThrAlaIleArg 452
RESULT 11
ABG06375
ID ABG06375 standard; Protein: 2570 AA.
XX AC ABC06375;
DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #6366.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX OS Homo sapiens.
XX PN WD200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS70562.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX PS Claim 20; SEQ ID No 36734; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostic products as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
```



CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 2570 AA:

Alignment Scores:  
 Pred. No.: 0.000223 Length: 2570  
 Score: 148.00 Matches: 133  
 Percent Similarity: 35.67% Conservative: 81  
 Best Local Similarity: 22.17% Mismatches: 195  
 Query Match: 3.87% Indels: 191  
 DB: 22 Gaps: 29

US-10-008-355-1 (1-2139) x ABG06375 (1-2570)

QY 491 CCAAAAAGAAATGCAGACAGAACCACTCTGCATCGTAGACCTTCTATTCCAACA 550  
 DB 1864 ProThrLysAsnValLysThrThrAsnCysPheSer  
 QY 551 ACGNATCTCCATCGCTACGATGATTAATCAAGGACGCTGATGGTATTGGCTCCTC 610  
 DB 1877 SerAsnThrArgLysMetThrSerLeuLeuGluLysThr  
 QY 611 CCAGCTCTGTAGTAAAGTTCGGAGCGCATACGACAACTGGATGTCGCCCGTCACACGG 670  
 DB 1889  
 QY 671 GCGACTTCAGCGTATTCGCGGTGTATCCCGGTGCGCGAACACCGCGCGGAATACA 727  
 DB 1890 SerLeuThrAsnTyrAlaThrSerLeuAsnThrProValSerTyrProTyrPro 1909  
 QY 728 ---GCAAGGACATAACCTATAAGCCGTTTACTTCGCTGCGGTATCCATCGAAGCT 784  
 DB 1910 SerSerAlaThrLeuProSerLeuThrSerPheValTyrSerProHisSerThrGluAla 1929  
 QY 785 ACAAGGCTGAGC-----ACTATCCATACCATCGCTGTTCCCGGCGAGTACGGATCGCT 838  
 DB 1930 GluLeSerThrProLysThrSerPro---ProThrSerGlnMetValGluPhe--- 1947  
 QY 839 ACCTCACTTCTGGGGTGTGGAAGATCGTATCGAAACAGAGAACATCCTCGTATCGAAG 898  
 DB 1948 ---ProValLeuGlyThrArgMetThrSerSerAsnThrGlnProLeuLeuMetThrSer 1966  
 QY 899 TTCGCGGTATCAACAGGATCTCGAGAGAGCCATGAGCG---CAGATCAGGCTACCC 955  
 DB 1967 TrpAsnIleProThrAlaGluGlySerGlnPheProIleSerThrThrIleAsnValPro 1986  
 QY 956 GTATCAAAATATGCCAGCAAGTCTCAGAGTGTCTAACTATTGGAATTCGATCGGTA 1015  
 DB 1987 ThrSerAsnGluMetGluThrGluThrLeuHisLeuValProGlyProLeu---SerThr 2005  
 QY 1016 TGAACCGCGTCTCGCTGCTTGTACGTGATAGTGTCTGTAAGCGTGCAGGAAAGACAT 1075  
 DB 2006 PheThrAla----- 2008  
 QY 1076 TCGCAGACTGGATCCGTAAGAACGGCAGAGTGTCTGTATGGCGATGATTGTCTTCTC 1135  
 DB 2009 SerGlnThrGly-----LeuVal 2014  
 QY 1136 TCGAAAAGGCTTATAAGGAGGACCAAGGCCAACCGTGTGAGTACTTATTGGACGAGA 1195  
 DB 2015 SerLysAspValMetAlaMetSerSerIlePro----- 2025

QY 1196 CGCTCTTCGTTGGTACCGAGGTGGTTCGTTTGCACAGTTTGGCAACGCTATGGCTACAA 1255  
 DB 2026 MetSerGlyLe-----LeuProAsnHisGlyLeuSer 2036  
 QY 1256 ATCTGTATGCTCATCGCGGTATCTCAATTCGCTTGACGACAAAGT-----ACAAG 1306  
 DB 2037 GluAsnProSerLeuSerThrSerLeuArgAlaIleThrSerThrLeuAlaAspValLys 2056  
 QY 1307 ACTACCTCCCTCGCTCGACCGTAAGCTGTCGCCGCCATCGCTCGATATTGTACGCCGCG 1366  
 DB 2057 HisThrPheGluLysMetThrThrSerValThrPro-----Gly 2069  
 QY 1367 GTATCCCTCGCGACAAAGCTCCCGCATATATCAAGAATGTAATCGACAGAAATTCAAG 1426  
 DB 2070 ThrThrLeuProSerIleLeu-----SerGly 2078  
 QY 1427 GCGACAGAAAGATATGCAGACTTCGTATTCGACAGAGTGTGTTCTTATATACGACA 1486  
 DB 2079 AlaThrSerGlySerValIleSerLysSerProIleLeuThrTrpLeuLeu----- 2095  
 QY 1487 AGTTCCATGCTGCTCAAGT---CCATGGACAAAGAAAGTTGCCAAGGCTATCGAGA 1543  
 DB 2096 SerSerLeuProSerGlySerProAlaThrValSerAsnAlaProHis---ValMet 2115  
 QY 1544 AGATCCGCGCAGTAGAGCTTTCCAAGAGCGTAATAGCTGCTGCTGCGCTATTCAGGCGC 1603  
 DB 2115 hrSerSerThrValGluValSerLysSerThrPheLeuThrSerAspMetIleSerAlaH 2135  
 QY 1604 ATCGGATGGCCATGCTATGCCATTCGAGAGGCGCAAGCGTC----- 1645  
 DB 2135 isPropheThrAsnLeu-ThrThrLeuProSerAlaThrMetSerThrIleLeuThrArg 2154  
 QY 1646 -----TTTTCTTTCGCGGTT 1660  
 DB 2155 ThrIleProThrProThrLeuGlyIleThrThrGlyPheProThrSerLeuProMet 2174  
 QY 1661 TGGGTGAGATGTACCCGCGAGTCTCTGCGAGCGATGCCAAGTTCACCATCGGTATGA 1720  
 DB 2175 SerIleAsnValThrAspAspIleVal-----TyrIleSerThrHisPro----- 2189  
 QY 1721 GCTACGGCTCCATCAAG-----GATATGAACCGCAGGACGGTG----- 1759  
 DB 2190 ---GluAlaSerSerArgThrIleThrAlaAsnProArgThrValSerHisProSer 2208  
 QY 1760 -----CCTGCTACAACTATCATACGACAGGCAAGGCGGTAT 1795  
 DB 2209 SerPheSerArgLysThrMetSerProSerThrThrAsp----- 2221  
 QY 1796 TGGAGAACGAGGATCTTAAGAGCGATGAGTTGCGGTACAGGAGAATATCCTCG----- 1849  
 DB 2222 ---HisThrLeuSerValGlyAlaMetProLeuProSerSerThrIleThrSerSerTrp 2240  
 QY 1850 -----ACCTCTTCGCGACCAAAACTATGCTGCTATGCTGCGAGACGCTCAGC 1897  
 DB 2241 AsnArgIleProThrAlaSerSerProSerThrLeuIleIleProLysProThrLeuAsp 2260  
 QY 1898 TCCATATCGGTTTCTTATCGAAACACACATCACGGCGGTAACTCCGCTAGCCCGCTAT 1957  
 DB 2261 SerLeu---LeuAsnIleMetThrThrSer-----ThrValProGlyAlaSerPhe 2277  
 QY 1958 TCGATAAAGACGCGCGTCTGTATGCTGCTTTCGATGGCAACTGGGAAGCTATGAGTG 2017  
 DB 2278 ProLeu-----IleSerThrGlyValThrTyrPro 2287  
 QY 2018 GTGACATCGAGT---TCGAACCCGATCTGCAGC---GCACAACTCAGCGTGCACATCGCT 2071  
 DB 2288 PheThrAlaThrValSerSerProIleSerSerPhePheGluThrThrTrpLeuAspSer 2307  
 QY 2072 ACGTCTCTTTCATGATGACAAATGGGTACGTACGCGCGCTCTCATCAAGAGC 2125  
 DB 2308 ThrProSerPheLeuSerThrGluAlaSerThrSerProThrAlaThrLysSer 2325





SQ Sequence 1795 AA;

Alignment Scores:		
Pred. No.:	0.000739	Length:
Score:	141.50	Matches:
Percent Similarity:	34.70%	Conservative:
Best Local Similarity:	19.33%	Mismatches:
Query Match:	3.70%	Indels:
DB:	22	Gaps:
	25	
		1795

US-10-008-355-1 (1-2139) x ABB69806 (1-1795)

Qy	91	CTCAACGAACTCAATCAGGAGAAATCTGGATCGAATCGGTGAGCTCGCGCTTTACGCTCCG	150
Db	414	LeuAsnAspIleAsnLysTyrGlnTyrLysArgTyr	425
Qy	151	TTGGATTTCGCTCTACAGTTTCGCAAGACCGCTCCATCCCAATGCCGTGGTTATCTTCGCT	210
Db	426	-----ThrTyrGlyThrAspLysAsnAspValThrGluAlaProGluIle-----Lys	441
Qy	211	GGCGGATGTAACCGGTATCAGAGTCCGATCAGGGCTGTATCTTTTACCACCAACCACTCC	270
Db	442	SerProLeuLysGlyLeuHisLeuSerGluAsnIleValIle-LeuProGluThrThrTh	461
Qy	271	GGATACGGTGCTATCCAGAGCCAAAGCAGCGTGATCAGCATATCTGCGCATGGTTTC	330
Db	461	rThrThrThrThr-----ThrLysProValValLeuThrCysProThrIleSe	478
Qy	331	GTTTTCGCGAGTGGTGAGGAGCTTCGATLTYCCGGGTCTTTCCGTGAAGTATCTCGCG	390
Db	478	r-----ProProAsp-----	481
Qy	391	AAGATCGTGAAGTAGCGGACGAGGAGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATG	450
Db	482	-----ThrThrProLysProSerThrThrThrAlaValThr-----	493
Qy	451	GAGCGTCTGCGCAAGCTCAGGAGGTATGCCAAAGAACTGGCCAAATA-----AAGAA	501
Db	494	-----LysSerThrProLysIleSerThrGlu	503
Qy	502	AATGCAGACGAGAACCAACTCTGCATCGTAGACCTTTCTATTCCAAACAACAAATATTC	561
Db	503	uGlnHisSerThrThrAlaLysThrThrThrLysArgProThrThrValThrGlu	523
Qy	562	CTCATCTGCTACGATGTATTCAAGACGGTTCGTATGGTATTTGCTCTCCAGCTCTGTGA	621
Db	523	uLysThrSerSer-----	527
Qy	622	GGTAAGTTCGGAGCGGATACGGACAACCTGGATGTGCGCGCGTCACA-----CGGCGCAC	675
Db	528	-----AlaThrGluLysProArgThrValValThrThrThrGlnLysArgSerThr	546
Qy	676	TTCAGCGTATTCGCGGTGATGCGGGTGGGACAAACGGCGCG-----CC	720
Db	546	rThrThrHisAsnThrSerProAspThrLysThrThrIleArgSerThrThrLeuSerPr	566
Qy	721	GAATACAGCAGGACAATAAACCCATAAGCCGTTTACTTCGCTGCCG-----TA	771
Db	566	oLysThrThrThrProSerThrThrProSerThrThrProSerThrThrThr	586
Qy	772	TCCATGCAAGGCTACAGGCTGACGACTATGCCA-----TGACCATCGGTTCCCG	822
Db	586	rProSerThrThrThrProSerThrThrThrProSerThrThrThrProSerThrThrTh	606
Qy	823	GGCAGTACGGATCGCTACCTCACTCTTGGGGGTGGGAAGATCGGTATPCGAAACAGCAAC	882
Db	606	rThrValLysValSerThrHisArgProArgThrThrSerGlnLysThrThrThrAlaSe	626
Qy	883	AATCCTCGTATCAAGTTCCGGGTATCAGCAAGGCATCTGGAAGGAACCCATGAGCGCA	942
Db	626	rThrThrThrLysLysThrThrThrSerProLysThrThr-----	639
Qy	943	GATCAGGCTACCGGTATCAAAATATGCCAGCAAGTATGCTCAGAGTCTCACTATTATGGAA	1007

Db	640	-----LysThrThrAspIleProThrSerThrThrSerLysLeuSerThrThrThr	656
Qy	1003	AATTTCGATGATGAACCGCGCTCGCTCGCTGCTGACGTGATAGTCTAAGCGTGC	1061
Db	656	rGlnLysThrThrThrHisLysPheThrAlaAlaThrThrSerThrGluLysPr	676
Qy	1063	GAGAAAGAGCATTCGCAGACTCGATCCGTAAAGACGGCAAGAGTCTGCTATGGCGAT	1122
Db	676	OLysThrThrThrGluLysThrSerThrValSerThrThr-----	689
Qy	1123	GTATTGCTTCTCTCGAAAGGCTTATAAGGAAGGACCAAGCCCAACCGTGAGATGACT	1188
Db	690	-----ThrLysLysSerThrGluSerSerProLysProThrSerSer-----	703
Qy	1183	TATTGACGAGACGCTCTTCGGTGGTACCG-----AGTGGTTCGTTTTCGA	1233
Db	704	-----ThrGlyLysProThrThrProLysProSerThrArgThrProThrThr	721
Qy	1231	CAGTTTCCCAACGATGGCTACAAATCCTGATGCTCATCGCGGTATCTCAAAATCGCTT	1299
Db	721	rThrLysValThrThrThrGlnIleThrThrThrProLeuArgSerSerThrGl	741
Qy	1291	GACGACAAGTACAAG-----ACTACCTCCCTCGCTCGACCGTAAGTGCTGCC	1341
Db	741	uThrThrSerThrGlnProProThrThrThrProGlnProThrThrThrThrThrLe	761
Qy	1342	GCCATGCTCGATATTGACCGCGGCTATCCCTGCCGACAAAGCTCCCCCATATATTCAAG	1401
Db	761	uThrValThrProLysThrSerThrThrThrThrThrGluLysProIleThrSer--	780
Qy	1402	AATGTAATCGACAAGAAATTCAAAGCGCACACGAAGAAGTATGACAGACTTCGTATTCGAC	1461
Db	781	-----SerProLysProThrThrThrThrGlnLysThrThrSerThrAlaProAsnTh	798
Qy	1462	AAGAGTGTGTTCTTATACGACAAGTTTCCATGCCATGCTCAAGTCCCATGGACAAGGA	1522
Db	798	rThr-----LysValAlaIleThrThrGlnLysGluThrThrProThrGlnSerTh	815
Qy	1522	AAGTTTCCCAAGGCTATCGAGAAGATCGCGCAGTAGACAGCTTTCCAAAGACGTAATAGCT	1581
Db	815	rSerThrThrIlePheThrArgLysThrThrThrAsnAsn-----	828
Qy	1582	GCTGCTCGCGCTATTACGGCCGATGCGATGCGCAATGCCTATGCCATTGAGAAGGGCAAG	1641
Db	829	-----ProGluProThrSerThrGluLysProIle	838
Qy	1642	CGTCTTTTCTTCCGGTTTTCGTGAGATGTACCCCGGACGTGCTCTGCCGACGATGCC	1701
Db	838	eThrSerThrThrProLysProSerThrThrThrProLysThrSerThrValAlaSerSe	858
Qy	1702	AAC TTCACCATGCGTATGACGTACGGCTCGATCAAGGGATATGAACCCGAGACGGTGC	1761
Db	858	rThrGluLysThrThr---IleSerSerProLysProThrThrGluLysSerThrGluAs	877
Qy	1762	TGGTACAACTATCATCGACAGCAGGCGTATTGAGAACGAGGACTCTAAGACGCGAT	1821
Db	877	nProThrThr-----AsnSerValLysThrSerAlaLeu-----	888
Qy	1822	GAGTTTCCGCTACAGGAAATATCCTCGACCTCTCCGACCA-----	1866
Db	888	uThrSerSerThrGlnArgAlaThrSerThrThrSerGluProThrLysThrThrGlnAs	908
Qy	1865	-----AAACATATGTCGTATGCGGAGAAGGTCACGCTCCATATCGTTCTTCATATC	1921
Db	908	nIleThrThrThrThrProLysProThrThrLeuLysThr---SerThrGlnGluAlaTh	927
Qy	1921	AACGACATCAGGGCGGTAACTCGGTAGCCCGTATTCGATAAGAACGGCGCTGATC	1981
Db	927	rThrSerThrGlnLysValSerThrValThrIleThrThrLysLysAlaThrGluSerSe	947
Qy	1981	GGTCTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGAT	2041

```
Db 947 rProLeuThrThrLeuSerThrGluGlu-----ProAsnThrThrProLysProLe 964
QY 2041 CTCAGCGCACATACGCGTGACATCGCTCTCTTCATGATTGACAAATGGGT 2100
Db 964 uArgThrThrProThrThrSerValThrAlaThrThrArgIleThrThrThrTh 984
QY 2101 CAGTGGCCCGCTCATCCAGAGC 2125
Db 984 rIleSerGluSerSerThrGluThr 992

RESULT 14
AAY31745
ID AAY31745 standard; Protein; 430 AA.
XX AC
XX AAY31745;
XX
XX 22-NOV-1999 (first entry)
XX
DE Mycobacterium tuberculosis specific DNA-encoded polypeptide.
XX
KW Tuberculosis; infection; diagnosis; DNA probe.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 4 /note= "encoded by TGA"
FT Misc-difference 6 /note= "encoded by TGA"
FT Misc-difference 20 /note= "encoded by TGA"
FT Misc-difference 29 /note= "encoded by TGA"
FT Misc-difference 54 /note= "encoded by TGA"
FT Misc-difference 64 /note= "encoded by TGA"
FT Misc-difference 69 /note= "encoded by TGA"
FT Misc-difference 89 /note= "encoded by TGA"
FT Misc-difference 99 /note= "encoded by TGA"
FT Misc-difference 114 /note= "encoded by TGA"
FT Misc-difference 119 /note= "encoded by TGA"
FT Misc-difference 129 /note= "encoded by TGA"
FT Misc-difference 159 /note= "encoded by TGA"
FT Misc-difference 169 /note= "encoded by TGA"
FT Misc-difference 182 /note= "encoded by TAG"
FT Misc-difference 185 /note= "encoded by TGA"
FT Misc-difference 219 /note= "encoded by TGA"
FT Misc-difference 259 /note= "encoded by TGA"
FT Misc-difference 269 /note= "encoded by TGA"
FT Misc-difference 291 /note= "encoded by TGA"
FT Misc-difference 323 /note= "encoded by TGA"
FT Misc-difference 339 /note= "encoded by TGA"
FT Misc-difference 349 /note= "encoded by TGA"
FT Misc-difference 356 /note= "encoded by TGA"
```

```
FT Misc-difference 366 /note= "encoded by TGA"
FT Misc-difference 410 /note= "encoded by TGA"
FT Misc-difference 421 /note= "encoded by TGA"
FT Misc-difference 421 /note= "encoded by TGA"
XX EP945462-A1.
XX 29-SEP-1999.
XX
XX 25-MAR-1998; 98EP-0302287.
XX 25-MAR-1998; 98EP-0302287.
XX (COUL ) CSIR COUNCIL SCI IND RES.
XX Kumar D, Srivastava BS, Srivastava R;
XX WPI; 1999-530042/45.
XX N-PSDB; AAX87940.
XX New nucleic acid molecules, useful for detecting and identifying
XX Mycobacterium tuberculosis
XX
XX Disclosure; Page 15-27; 42pp; English.
XX
XX The present sequence represents an amino acid sequence deduced
XX from the Mycobacterium tuberculosis specific DNA fragment provided
XX in AAX87940. This DNA fragment comprises a StuI-StuI fragment of
XX M. tuberculosis genomic DNA and contains an insertion sequence-like
XX element and repetitive sequences. The DNA fragment is useful as a
XX probe, especially for detecting or identifying M. tuberculosis in
XX clinical isolates and body fluids e.g. sputum, cerebrospinal fluid,
XX pleural fluid, urine, gastric lavage, bronchial lavage, pericardial
XX or lymph node aspirate (all claimed). It is also useful for
XX restriction fragment length polymorphism analysis of M. tuberculosis
XX isolates (claimed). The probe provides rapid and specific diagnosis
XX of tuberculosis and M. tuberculosis infection.
XX
XX Sequence 430 AA;
SQ
Alignment Scores:
Pred. No.: 0.000595 Length: 430
Score: 138.00 Matches: 133
Percent Similarity: 23.61% Conservative: 28
Best Local Similarity: 19.50% Mismatches: 169
Query Match: 3.49% Indels: 352
DB: 20 Gaps: 37
US-10-008-355-1 (1-2139) x AAY31745 (1-430)
QY 1960 CGAATACGGGGCTACCGAGTTACCGC-----CCGTGATGTCGTTGTCATAGGAAG 1907
Db 10 ArgArgArgSerLeuArgCysValArgGln***ProSerAlaArgCysLeuTrpGly*** 29
QY 1906 CGATATGGAGCTGACCGTCTTCGCGCATAGCGACCATAGTTTGGTGGCGAAGAGGTCGA 1847
Db 30 IleGlySerGlyAspArgGluIleAspGlyGlu-----SerGlyPro 43
QY 1846 GGATATTCTCTGTACGGCAACTCATTCGCTCTTAGGATCCTGCTTCCAATACGCCCT 1787
Db 44 GlyTyrArgProProArgSerHis-----Cys***SerArgPro 57
QY 1786 TGCCTGTCG-----TATGATAGTTGTACCGGACCGCTCCTCGGTTTCATATCCCTTGA 1733
Db 58 CysArgLysArgArgTyr***ArgCysArgAsp***ArgSerArgCys----- 73
QY 1732 TGGAGCCGTAGCTCATACGCATGCTGAAGTTGGCATCGCTCGGACGACGACGTCGGGGT 1673
Db 74 TrpCysHisProGlyTyrSerCysArgTyrTrpCys-----HisProArg*** 89
QY 1672 ACA----- 1670
```

```

Db 90 SerCysArgCysTrpCysTyrArgArg***SerArgArgCysCysSerCysProCysCys 109
Qy 1669 ---TCTCAGCAACCGGCAAGAAAGACCGCTTGCCTTCTCAATGGCATAGGCAT--- 1616
Db 110 ArgSerProCys***SerArgArgArg***GlyCysProCysCysSerCysGlnHis*** 129
Qy 1615 -----TGGCCATCGCATCGCCTGTAATAGCGC 1589
Db 130 GlyCysArgTyrCysArgTyrProGlySerArgTyrProSerSerArg----- 145
Qy 1588 GAGCAGCAGCTATTACGCTTTGTGAAAGCTCTACTGCGGATCTTTCTCGATAGCCTTGG 1529
Db 146 -----CysProSerLeuArg-----Cys 151
Qy 1528 CAACACTTTCTCTGTCCA-----TGGACTTGAGCATGGCATGGA 1490
Db 152 ArgArgPheArgCysProArg***ArgCysGlnArgTyrTrpCysProAsn***ThrGly 171
Qy 1489 ACTTGT-----CGCTATAAGGAACACACACTCTTGTGCA 1457
Db 172 ArgCysCysArgCysProSerSerArg***GlnTyr***SerProAlaGlyCysArg 191
Qy 1456 ATACGAAGCTCTGCATCTCTCTGCTGCGCTTTGAATTTCTTGTCTGCATACATTTCTGA 1397
Db 192 -----ArgThrAlaArgCysArg-----CysCys----- 199
Qy 1396 ATATATCGGGGAGCTTCGCGCAGGATACGCCGCGCTACAAATATCAGCATGGCGGCA 1337
Db 200 -----CysCys-----ArgCysTrpArgThr 206
Qy 1336 GCACCTTACGCTGAGCGAGGGAGGTAGTCTTTGTACTTGTCTCAAGCATTTGAGGA 1277
Db 207 -----ArgCysCysCysCysArg----- 212
Qy 1276 TACCGGATGAGCATCAGGATTTTACCCAAATGCGTTGGCAACTGCGAAACAGACCA 1217
Db 213 -----CysTrpGlnSerLeuGly***SerArg 221
Qy 1216 CTTGGTACCACCGAGAGCGTCTCGCTCAAAATAGTCATCTACGCTTGGCCTTGGCTC 1157
Db 222 ProArg-----SerArgSerArgCysSerArg----- 231
Qy 1156 CTTCTTTAAGCCTTTTCGAGAGAAGACAATACATCGCCATAGACAGCACTCTTGGCGT 1097
Db 232 -----ArgArgPheGlnAsnArgCysCysArg 240
Qy 1096 TCTTACGGATCCAGCTCTGCGAATGCTC---TTTCCTCGGCACGCTTACGACCTATCAGT 1040
Db 241 SerArgGlyPheArgIleArgCysCysSerPhePro----- 252
Qy 1039 CAAGCAGCGAGAGCGGGTTTCAATACCGATCGAATTTCTTCCAATAGTTAGCACTCTGAG 980
Db 253 -----GlyPheArgAsnArg-----His***Ile 260
Qy 979 CATACTTGTGGCATATTTGATACGGGTAGCCTGATCTGCGCTCATGCTTCTCTCCAGA 920
Db 261 LeuArgCysPheHis----- 265
Qy 919 TGCCTTCTGTGATACCGGAACCTTCGATACGAGGATTTGTTCTCGTTTTCGATACGATCTT 860
Db 266 -----CysArgTyr***SerCysArgArgCysArg----- 275
Qy 859 CCACACCCCAAGAAGTAGGTAGGATCCCGTACTGCGCGGGAACCGCATGCTCATGGCAT 800
Db 275 ----- 275
Qy 799 AGTCGTAGCCTTTAGCCTTTGCATGATACGGCAGCGAAAGTAAACGGGCTTATAGGCTT 740
Db 275 ----- 275
Qy 739 TATTGTCTCTGTATTGCGCGCGCGGTGTCGGCACCGGCATACACGGCGGAATACGC 680
||||| ||| |||||||

```

```

Db 276 ---CysProArgCysPheGlyCysArgGlyCysArg----- 286
Qy 679 TGAAGTCGCGCTGTGACGCGGCACATCCAGTTGTCGTTATCGCCTCCGAACCTTACCTA 620
Db 287 -----CysGlnGlyCys***SerHisArgArgPheArgCysArgGluCysCys 302
Qy 619 CAGAGCTGGGAGGAGCAATACCATACGAACGCTCTTGAATACATCGTAGACGATGAGGA 560
Db 302 ----- 302
Qy 559 AGTATTCTGTTGGAATAGAAAGGCTCTACGATGCGACAGTTGTTCTGCTCTGCATTTT 500
Db 303 AsnCysArgCysTrp-----ArgCysArgGluCysSerArgArgProGly 317
Qy 499 CTT----- 497
Db 318 LeuProGlyArgAsp***ArgProValGlyHisArgLysIleProThrCysCysPheArg 337
Qy 496 -----TTTGGCCAGTTCTTGGCATACCTCTCTGAGCTTTTGGCAGAC 455
Db 338 Cys***ArgSerProArgSerArgProAlaLeu***TrpProGlySerCys***Thr 357
Qy 454 GCTCCATCTCGTCAGTCATACCCCTTGAGCTGCTCTTCTACCTTCTCGTCTGATACACCG 395
Db 358 Asn-----ProIleArg 361
Qy 394 TCTTGGCAGACTTCTACGGAAGACCGGAATCGGAAGCTCCTCACCCATCGTGCAG 335
Db 362 CysCysProSer***SerArgProIlePro-----AlaArgProArgLeuPro 377
Qy 334 AAACGAACCATCGCGCAGATAGTCGTGATCCACCGCTTGGCTCTGATACACCGT 275
Db 378 GlyArgSerTyrArgTrp-----ProPro----- 385
Qy 274 ATCCGAGTGGTGT-----TGTAAAGATCAGCCCTGATCGGACACTGTGA 227
Db 386 ThrLysSerGlyArgSerGlnAsnCysTrpHisArgSerGlySerArgThr----- 403
Qy 226 TACCGGTACATCCGCCACCGA-----AGATAACCGGCAATGGCAATGGACGGCTGT 173
Db 404 ---ArgTyr---ArgHisArgCys***ArgIleProThrAlaHisTrpSerSer*** 421
Qy 172 CGAAAC 167
Db 422 ArgSer 423

```

```

RESULT 15
ABB61705
ID ABB61705 standard; Protein; 684 AA.
XX ABB61705;
AC AC
XX XX
DT 26-MAR-2002 (first entry)
XX XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11907.
XX XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX XX
OS Drosophila melanogaster.
XX XX
PN WO200171042-A2.
XX XX
PD 27-SEP-2001.
XX XX
PF 23-MAR-2001; 2001WO-US09231.
XX XX
PR 23-MAR-2000; 2000US-191637P.
XX XX
PA 11-JUL-2000; 2000US-0614150.
XX XX
PI (PEKE ) PE CORP NY.
Venter JC, Adams M, Li PWD, Myers EW;

```



XX WPI: 2001-656860/75.  
 DR N-PSDB; ABL05808.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 11907; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB57737-AB572072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 684 AA;

Alignment Scores:  
 Pred. No.: 0.000823 Length: 684  
 Score: 138.00 Matches: 151  
 Percent Similarity: 34.85% Conservative: 109  
 Best Local Similarity: 20.24% Mismatches: 264  
 Query Match: 3.49% Indels: 222  
 DB: 22 Gaps: 36

US-10-008-355-1 (1-2139) x ABB61705 (1-684)

QY 2067 GATGTCACCGCTGATGCTGGCTGCAGATCGGTTGCAACTCGATGTCACCATCATAGC 2008  
 Db 3 AspLeuHisValAspGluLeuValValAspLeuGluValValMetLeuHisArg 22  
 QY 2007 TTCCCATGTTCCCATCAAGAACAGACGATCAGACGGCCGCTTATATGCAATACGGGGCT 1948  
 Db 23 -----ProValGluGlyGluThrGluGlyGlnGlyValHisValAlaGlyThrVal 39  
 QY 1947 ACCGGAGTTACCGCCGCTGATGCTGTTGTCGATAGGAAGCGATATGAGCTGACCGCTT 1888  
 Db 40 SerHisAlaAlaAspHisArgIleValAla-----SerGluLeuGlyAsn 55  
 QY 1887 CTGGCATACCGACCATAGTTTTTGGTGGCGAAGAGCTCGAGGATATT-----CTCCTG 1834  
 Db 56 LeuGlyLeuValValArgValAspHisValGluGluLeuGluHisLeuAlaLeuLeu 75  
 QY 1833 TACGGCAAA-----CTCATCGCTCTTAGGATCCTGCTTCTC 1798  
 Db 76 AlaAlaArgGlnHisHisSerValGlyHisAlaAlaMetGlyLeuArgLeuIleGlnLeu 95  
 QY 1797 CAATACCCCTTCGCTGCTGATGATGTTGTCGATAGGATGACAGGACCGTTC-----CTCGG 1747  
 Db 96 Glu-----LeuValValThrValGlyValHisGlnLeuVal 109  
 QY 1746 TTCAATATCCTTGATGAGCGGATGATGATACGATGATGATGATGATGATGATGATGATGATG 1687  
 Db 110 GluLeuGlnValGlyAlaIleValHisProGluGlyVal-----124  
 QY 1686 AGCAGCTCCGG-----GTACATCTCAGCGAAACC 1657  
 Db 125 ---ThrAlaGlyAlaLeuAspValValAsnAlaArgLeuGlnMetHisLeuGluValLeu 143  
 QY 1656 GCGAAGAAAGAGCTTGCCCTTCTC-----AATGGCATAGCATGCGCATCGC 1606  
 Db 144 GlnLeuGluGluValLeuValValValLeuGlnTrpAspLeuValSerValLeuIleVal 163  
 QY 1605 ATCGGCTGAATAGCGGACGACGATATACGCTCTTGGAAAGCTCTACTGCCGATC 1546  
 Db 164 LeuGlySerAsnGluAspLeuHisAsn-----172

QY 1545 TTTCTCGATAGCCTTGGCAAACTTTTCTTCTCATGCACTTGAGCATGGCATGAACCTT 1486  
 Db 173 -----GlyIleLeuSerLeuValArgGlyHisValGlnGly---GluLeu 186  
 QY 1485 GTCGCTATAGGAAACACACTCTTGTGCAATACGAACCTGTCATCTTCTGTCGTCGCC 1426  
 Db 187 GluGlyLeu-----ValValGluThrLeuLeuAlaGlyLeuAspHisIleAla 202  
 QY 1425 TTTGAATTTCTGTGATTACATTTCTTCAATATATC-----GGGAGCTTGTCTC 1378  
 Db 203 GlySerLeuValValGlyValProAlaGluGlyValGlnArgIleGlyGlySerLeuAsn 222  
 QY 1377 GGC-----AGGATACCGCCGCGGTACAAATATCATGACATGCGCGG 1339  
 Db 223 GlyGlyLeuHisIleAspGlyIleAlaAspValArgValGluValLeuGluValGlyGly 242  
 QY 1338 CAG-----CACCTTACGGTTCGAGGAGGAGGAGGTAGTCTTTGTACTT 1297  
 Db 243 GlnThrPheGlyHisValAspGlyIleAspValAspThrLeuGluGlyValIleLeuLeu 262  
 QY 1296 GTCGTCAAGCGATT-----GAGGATACCGCGCATGAGCATC 1261  
 Db 263 ValValLeuArgIleGlyLeuHisGlnMetProValValAlaAspAsnGlyLeuValTyr 282  
 QY 1260 AGGATTTGTAGCCAATCGTTGGCAAACTGTGCAAAACAGAACCACTCGGTACCCCGAA 1201  
 Db 283 LeuValGluAspArgIleAla-----HisArgGlyValValVal 295  
 QY 1200 GAGGCTCTCGCTCAATAAGTCATCTCACGGTGGCGCTTGGCTCTTCTTATAAGCTT 1141  
 Db 296 HisGluLysAlaArgGluAlaLeuValThrMetAlaIle-----MetValAlaLeu 312  
 QY 1140 TTCGAGAGAAGCAATACATCGCATAGACAGACACTCTTCCGCTTCTTACGGATCCAGTC 1081  
 Db 313 ValHisAsnGluAlaValGlnLeuValValThrAlaValGluLeuLeuValHisAlaVal 332  
 QY 1080 TGGCAATGCTCTTTCTCTCGGCACGCTTACGACCTATCAGCTCAAGACGAGCGACGCG 1021  
 Db 333 -----ArgIleAlaLeuAspAspValAla 340  
 QY 1020 GTTCATACCATCGAATTTCTCCATAGTATTAGTACTCTGACATACTTGTGGCATATTT 961  
 Db 341 GlnHisIleAspGlyLeu-----HisIleGluLeuValLeuValAspProValHis 358  
 QY 960 GATACGGGTAGCCTGATCTCGCTCATGCTCTCTTCCAGATGCTTGTGATACCGCG 901  
 Db 359 AspThr-----LeuArgValHisLeuHisMetLeuAspValLeuVal-----372  
 QY 900 AACTTCGATACGAGGATTTCTCGTTTTTCGATACGATCTTCCACACCCCAAGAGTGA 841  
 Db 373 -----ValValLeuValValGluAlaValValIleLeuPro-----384  
 QY 840 GTACCGATCGTACTGCCCGGGAACCGATGGTTCATGCGCATAGTCTGTCAGCTTGTAGCC 781  
 Db 385 ValProLeu-----GlyTyrAlaValVal-----392  
 QY 780 TTGCATGATACGCGACGAGGAAGTAAACGGCTTATAGGTTTATGCTTCTTGTGTATTC 721  
 Db 393 -----GlyAlaGlnValThrGlyIleAlaGlyPhe-----402  
 QY 720 GGCGCGCGGTTGTGCGCACCGGCATACACGCGGAATACGCTGAAGTCGCCCGTGTG---664  
 Db 403 -----GlyTyrPheHisIleHisValGluSerGlyHisLeuAlaGluIleVal 417  
 QY 663 -----ACGGCCACATCCCACTTGTCTCGTATCGCTCGGCACTT 625  
 Db 418 GlyGlnThrLeuGlnValProThrGlyGlnLeuProValAspIleGlnValGluLeu 437  
 QY 624 ACCTACAGACTGGGAGGAGCAAAATACCATACGACCTCTTGAATACATCGTAGAGCAT 565  
 Db 438 Thr-----AlaThrIleLeuValLysAsp 445

```
Oy 564 GAGGAAGTATTCGTTGTTGGAATAGAAAGGCTCTACGATCGAGAGTTGGTTCTCGTCTGC 505
Db   :::      :::      |||||:::      |||||:::      |||||:::
446 MetLeuLeuAlaAlaLeuArgProGluArgValGlnIleGluValValValGlu--- 464
Oy 504 ATTTTCTTTTGGCCAGTTCTTGGCATACCTCTGAGCTTTGGCCAGACGCTCCATCTC 445
Db   |||      |||      |||      |||      |||      |||
465 -----GlyIleGluAlaAsnIleLeuGlnValValGlyGlnIleValIleVal 480
Oy 444 GTCAGTGATACCGCTTGAGCTGCTCTTACCTGCTCGTTACCTTACGATCTTTGGCGCAG 385
Db   |||      |||      |||      |||      |||      |||
481 ValGlnGlyThrValHisValArg-----IleValArgThrAspHisHisValValAsp 498
Oy 384 ATACTTCACGGAAGACCCGGAATCGGAAGCTCTCACCCATCGTCGAGAAACGAAACC 325
Db   :::::||||
499 ValValHis-----Leu 502
Oy 324 ATCGCGCAGATAGTCGTATCCAC-----CGTGCTTTGGCTCTG 286
Db   :::      :::      |||||:::      |||||:::
503 LeuAspMetMetGluAlaValHisPheValValLeuSerProPheGlnCysLeuAlaIle 522
Oy 285 GATAGCACCC-----GTATCCGCGAGTGGTAAAGATCAGGCCCTGATCGGACAC 232
Db   :::      |||      |||||      |||||      |||||      |||||
523 GluValAlaProTPrpValAspLeuValValGlyArgGlnAspAspSerLeu----- 539
Oy 231 TGTGATACCGGTATCCGCCACCGCAGATAAACCAGCGCATTTGGCAATGGACGGCTTGTG 172
Db   |||      |||      |||      |||      |||      |||
540 ---HisIleIleThrGluGlnThrGlyAsnGlyHisGlyValHisGluGlnLeuVal 558
Oy 171 -----GAAACTGTACAGCGAATCCCAACGGGACGTAAGGCC 136
Db   :::      |||      |||      |||      |||      |||
559 TyrValLeuAlaGlyProValAspValLeuGluGlyLeuProValLeuGlyGluVal 578
Oy 135 GAGCTCACGCATTCGATCCAGATTCTCTGATTGAGTTGCTTT-----GAGGAGCCACAT 82
Db   |||      |||||      |||||      |||||      |||||      |||||
579 GluGluThrHisGlnThrHisHisIleLeuLeuValLeuValHisHisArgProHis 598
Oy 81 GCCTTTGTCGGCTTTGGC 64
Db   :::::||||
599 HisValIleGlyGluGly 604
```

Search completed: May 16, 2003, 12:28:32  
Job time : 165 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:23:56 ; Search time 30.5 seconds  
(without alignments)  
4126.927 Million cell updates/sec

Title: US-10-008-355-1  
Perfect score: 3820  
Sequence: 1 atgcaaatgaattaaaaag.....aagagctgaagtgtgatctaa 2139

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USTO.spool/US10008355/runat\_16052003\_110339\_9386/app\_query.fasta\_1.2311  
-DB=Issued\_Patents\_AA -OFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355 @CGN\_1.1.33 @runat\_16052003\_110339\_9386 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NSG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA:  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	138	3.5	430	4	US-09-156-836B-2
c 2	130	3.4	883	2	US-08-953-492-2
c 3	129	3.3	430	3	US-08-997-897-2
c 4	127.5	3.3	1194	2	US-08-488-940-18
c 5	123.5	3.2	713	4	US-09-059-584-53
c 6	122.5	3.2	800	2	US-08-488-940-3
c 7	122.5	3.2	813	2	US-08-488-940-3
c 8	122.5	3.2	1181	2	US-08-488-940-2
c 9	122.5	3.2	1194	2	US-08-488-940-1
c 10	122.5	3.2	1194	2	US-08-488-940-17
c 11	119	3.0	2152	4	US-09-036-987A-3
c 12	119	3.0	2152	4	US-09-370-700-3

ALIGNMENTS

RESULT 1

US-09-156-836B-2

Sequence 2, Application US/09156836B

Patent No. 6242585

GENERAL INFORMATION:

APPLICANT: Srivastava, Ranjana

APPLICANT: Kumar, Deepak

APPLICANT: Srivastava, Brahm Shanker

TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT

FILE REFERENCE: U 011876-4

CURRENT APPLICATION NUMBER: US/09/156.836B

CURRENT FILING DATE: 1998-09-18

PRIOR APPLICATION NUMBER: 08/997.897

PRIOR FILING DATE: 1997-12-24

NUMBER OF SEQ ID NOS: 7

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 430

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

FEATURE:

NAME/KEY: UNSURE

LOCATION: (4)

OTHER INFORMATION: amino acid has not been identified

NAME/KEY: UNSURE

LOCATION: (6)

OTHER INFORMATION: amino acid has not been identified

NAME/KEY: UNSURE

LOCATION: (20)

OTHER INFORMATION: amino acid has not been identified

NAME/KEY: UNSURE

LOCATION: (29)

OTHER INFORMATION: amino acid has not been identified

NAME/KEY: UNSURE

LOCATION: (54)

OTHER INFORMATION: amino acid has not been identified

Sequence 33, Appl  
Sequence 33, Appl  
Sequence 14, Appl  
Sequence 14, Appl  
Sequence 6, Appl  
Sequence 6, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 23, Appl  
Sequence 3032, Ap  
Sequence 4, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 511, Ap  
Sequence 19, Appl  
Sequence 30, Appl  
Sequence 34, Appl  
Sequence 40, Appl  
Patent No. 5223423  
Sequence 2, Appl  
Sequence 28, Appl  
Sequence 295, App  
Sequence 6, Appl  
Patent No. 5200183  
Sequence 9, Appl

```

; NAME/KEY: UNSURE
; LOCATION: (64)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (69)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (89)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (99)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (114)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (119)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (129)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (159)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (169)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (182)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (185)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (219)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (259)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (269)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (291)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (323)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (339)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (349)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (356)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (366)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (410)
; OTHER INFORMATION: amino acid has not been identified
; NAME/KEY: UNSURE
; LOCATION: (421)
; OTHER INFORMATION: amino acid has not been identified
; US-09-156-836B-2

```

```

Alignment Scores:
Pred. No.: 7.39e-05 Length: 430
Score: 138.00 Matches: 133
Percent Similarity: 23.61% Conservative: 28
Best Local Similarity: 19.50% Mismatches: 169

```

```

Query Match: 3.49% Indels: 352
DB: 4 Gaps: 37
US-10-008-355-1 (1-2139) x US-09-156-836B-2 (1-430)
Qy 1960 CGAATACGGGGCTACCGGAGTTACGCG-----CCGTGATGTCGTTGTCGATGAGGAAG 1907
Db 10 ArgArgArgSerLeuArgCysValArgGln***ProSerAlaArgCysLeuTrpGly*** 29
Qy 1906 CGATATGGAGCTGACCGTTCTCGGCATAGCAGCACCATAGTCTTTTGGTCGGAAGAGGTGCA 1847
Db 30 IleGlySerGlyAspArgGluIleAspGlyGlu-----SerGlyPro 43
Qy 1846 GGATATTCTCTGTACGGCAAACTCATCGCTCTTAGGATCCTGCTTCTCCAAATACGCCCT 1787
Db 44 GlyTyrArgProProArgSerHis-----Cys***SerArgPro 57
Qy 1786 TGCCTGTGCG-----TATGATAGTTGTACAGGCACCGCTCTCGGGTTCATATCCCTTGA 1733
Db 58 CysArgLysArgArgTyr***ArgCysArgAsp***ArgSerArgCys----- 73
Qy 1732 TGGAGCCGTAGCTCATACGCATGCTGAAGTTGGCATCCCTCGGCAGACGACGTCGGGGT 1673
Db 74 TrpCysHisProGlyTyrSerCysArgTyrTrpCys-----HisProArg*** 89
Qy 1672 ACA-----TCGCCATCGCATCGCCCTGAATAGCGC 1670
Db 90 SerCysArgCysTrpCysTyrArgArg***SerArgArgCysCysSerCysProcysCys 109
Qy 1669 ---TCTACGCAACCGGCAAGAAAGACGCTTTCGCTTCTCAATGGCATAGGCAT--- 1616
Db 110 ArgSerProCys***SerArgArgArg***GlyCysProCysCysSerCysGlnHis*** 129
Qy 1615 -----TCGCCATCGCATCGCCCTGAATAGCGC 1589
Db 130 GlyCysArgTyrCysArgTyrProGlySerArgTyrProSerSerArg----- 145
Qy 1588 GAGCAGCAGCTATTACGCTCTTGGAAAGCTACTCGCGGATCTTCTCGATAGCCTTGG 1529
Db 146 -----CysProSerLeuArg-----Cys 151
Qy 1528 CAACACTTTTCTCTGTCCA-----TGGACTTGAGCATGCGATGGA 1490
Db 152 ArgArgPheArgCysProArg***ArgCysGlnArgTyrTrpCysProAsn***ThrGly 171
Qy 1489 ACTTGT-----CGCTATAAGGAACACACACTCTTGTCGA 1457
Db 172 ArgCysCysArgCysProSerSerArg***GlnTyr***SerProAlaGlyCysArg 191
Qy 1456 ATACGAAGTCTGCATCTTCTTCGTGCGCTTTTGAATTTCTTCTCGATTACATCTTGA 1397
Db 192 -----ArgThrAlaArgCysArg-----CysCys----- 199
Qy 1396 ATATATCGGGGAGCTTGTGCGCAGGGGATACCCCGCGGTACAAATATCGAGCATGCGGGCA 1337
Db 200 -----CysCys-----ArgCysTrpArgThr 206
Qy 1336 GCACCTTACGCTCGAGCGAGGGAGGTAGTCTTTGTTACTTGTCTGCTCAAGCGATTGAGGA 1277
Db 207 -----ArgCysCysCysArg----- 212
Qy 1276 TACCGGCATGAGCATCAGGATTTGTAGCCAATGCGTTGGCAAACTGTGCAAAACGAACCA 1217
Db 213 -----CysTrpGlnSerLeuGly***SerArg 221
Qy 1216 CCTCGGTACCACCAAGAGCGTCTCGCTCAAAATAGTCATCTACGGTTGGCCTTGGGTC 1157
Db 222 ProArg-----SerArgSerArgArgCysSerArg----- 231
Qy 1156 CTTCTTTAAGCCTTTTCGAGAGAAGACAATACATACCCATACAGCAGCACTCTTTCGCGT 1097
Db 232 -----ArgArgPheGlnAsnArgCysArg 240

```

```
QY 1096 TCTTACGATCCAGTCGTGCGAATGCTC---TTTCTCGGCACGCTTACGACCTATACAGT 1040
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 241 SerArgGlyPheArgGlyLeuArgCysSerPhePro----- 252
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 1039 CAAGACGAGCAGACCGCGTTTCATCCGATCGAATTCCTCCATATAGTTAGCACTCTGAG 980
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 253 ---GlyPheArgAsnArg-----His***Ile 260
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 979 CATACTTCTGCGCATATTTGATAGCGGTAGCCTGATCGCGCTCATGGCTTCCTTCCAGA 920
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 261 LeuArgCysPheHis----- 265
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 919 TGCCTTCTGATACCGGAACTTCGATACGAGATTGTTCTCTCTTTTCGATACGATCTT 860
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 266 ---CysArgTyr***SerCysArgArgCysArg----- 275
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 859 CCACACCCCAAGAGTAGGTAGCGATCCGTACTCTCCCGGGAACCGCATGTCATGCGCAT 800
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 275 ----- 275
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 799 AGTCGTAGCCTTGTAGCCTTGTGATGATACGGCAGCGAAGTAACGGCTTATAGGTT 740
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 275 ----- 275
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 739 TATTGCTCTGCTTATTCGCGCGCGGTTGTCGGCACCGGCATACACGCGGAATACGC 680
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 276 ---CysProArgCysPheGlyCysArg----- 286
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 679 TGAAGTCGCGCGTGTGACGGCGGCACATCCAGTTGTCGGTATCGCTCCGNACTTACCTA 620
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 287 ---CysGlnGlyCys***SerHisArgArgPheArgCysArgGluCysCys 302
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 619 CAGAGCTGGGAGGAGCAATACCATACGAGCTCTTGAATACATCCTAGACGATGAGGA 560
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 302 ----- 302
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 559 AGTATTGTTGTTGGATAGAAAGCTCTACGATGCGAGTTGTTCTCGTCTGCATTTT 500
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 303 AsnCysArgCysTrp-----ArgCysArgGluCysSerArgArgProGly 317
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 499 CTT----- 497
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 318 LeuProGlyArgAsp***ArgProValGlyHisArgLysIleProThrCysCysPheArg 337
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 496 -----TTTGGCCAGTTCTTGGCATACCTCTCGCATCTTGGCGAGAC 455
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 338 Cys***ArgSerProArgSerArgProAlaLeu***TrpProGlySerCys***Thr 357
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 454 GCTCCATCTCGTCAGTATACCCCTTGAGCTGCTCTTACCTTGTCTGCTTACCTTCAGGA 395
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 358 Asn-----ProIleArg 361
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 394 TCTTGGCAGATCTTACGGAAGAACCGGAATCGGAAGCTCCTCACCCATCGTGGAG 335
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 362 CysCysProSer***SerArgProIlePro-----AlaArgProArgLeuPro 377
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 334 AAACGAAACCATCCGCGAGATGATGATCCACCGCTTGTGCTCTGGATAGCACCGT 275
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 378 GlyArgSerTyrArgTrp-----ProPro----- 385
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 274 ATCCGCGAGTGTGT-----TGTAAGATACGAGCCCTGATCGGACACTGTGA 227
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 386 ThrLysSerGlyArgSerGlnAsnCysTrpHisArgSerGlySerArgThr----- 403
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 226 TACCGGTACATCCGCCACCGA-----AGATAACACCGCATTTGGCAATGACGCGTGT 173
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 404 ---ArgTyr---ArgHisArgCys***ArgIleProThrThrAlaHisTrpSerSer*** 421
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 172 CGAAAC 167
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
QY 422 ArgSer 423
```

RESULT 2

```
US-08-953-492-2
; Sequence 2, Application US/08953492
; Patent No. 5849555
; GENERAL INFORMATION:
; APPLICANT: Brown, James
; APPLICANT: Jaworski, Deborah
; APPLICANT: Lawlor, Elizabeth
; APPLICANT: Wang, Min
; TITLE OF INVENTION: NOVEL vals
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,492
; FILING DATE: 17-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/844,064
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9607991.8
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31458-4/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 883 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-953-492-2
```

```
Alignment Scores:
Pred. No.: 0.000829 Length: 883
Score: 130.00 Matches: 104
Percent Similarity: 32.68% Conservative: 63
Best Local Similarity: 20.35% Mismatches: 168
Query Match: 3.40% Indels: 176
DB: 2 Gaps: 24
```

US-10-008-355-1 (1-2139) x US-08-953-492-2 (1-883)

```
QY 574 CATGATATTC-----AAGGACGTTGTTATGTTGCTCTCT 609
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 26 AspValPheLysProSerGlyAspGlnLysAlaLysProTyrSerIleValIlePro 45
QY 610 CCCAGCTCTGTAGTAGTTCGAGCGGATACGGACAACTGGATGTGCGCGGTACACAG 669
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 46 ProAsnValThrGlyLysLeu-----HisLeu 54
QY 670 GCGCACTTCAGCGTATTCGCGGTGTATCGCGTGTATCGCGTGTATCGCGTGTATCGAG 729
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
Db 55 Gly-----Hisala 57
QY 730 AAGGACAAATAAACCCCTATAAGCCGCTTTACTTTCGCTGCCGTATCCATCGAAGGTACAG 789
Db 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
```

Db 58 TrpAspThr---ThrLeuGlnAspIleIleIleIleArgGlnLysArgMetGlnGlyPheAsp 76  
QY 790 GCTGACGACTATGCCATGACCATCGTGTTCCTCCGCGCAGTACGATCGCTACCTC---ACT 846  
Db 77 -----ThrLeuTrpLeuProGlyMetAspHisAlaGlyIleAlaThr 90  
QY 847 TCTTGGGGTGTGAAGATCGTATCGAAAAACGAGAACAAATCCTCGTATCGAATTCGCGGT 906  
Db 91 GlnAlaLysValGluGluArgLeuArgGlyGluGlyIleSerArgTyrAspLeuGlyArg 110  
QY 907 -----ATCAAGCAAGGCATCTGGAAGAGAGCCATGAGCGCAGATCAGGCTACCC 954  
Db 111 GluSerPheLeuThrLysValTrpGluTrpLysAsp-----GluTrpAlaThr 126  
QY 955 CGTATCAAAATATGCC---AGCAAGTATGCTCAGAGTCTAACTATTGGAAGAAATTCGATC 1011  
Db 127 ThrIleLysGluGlnTrpGlyLysMetGlyLeuSerValAspTyrSerArgGluArgPhe 146  
QY 1012 GGTATGAACCGCGGTCTCGCTCTCTTGACGTGATAGGTCTGAAGCGTCCCGAGGAAAGA 1071  
Db 147 ThrLeuAspGluGlyLeuSer-----LysAlaValArgLys 158  
QY 1072 GCATTCGCAGACTGATCGCTGAAG-----AACGGCAAGAGTCTCTCTATGCGC 1119  
Db 159 ValPheValAspLeuTyrLysGlyTrpIleTyrArgGlyLysPheIleIleAsnTrp 178  
QY 1120 GAT-----GTATTGTCTCTCGAAAAGGCTTATAAGGAAGAGCCCAAG 1164  
Db 179 AspProAlaAlaArgThrAlaLeuSerAspIleGluValIleHisLysAspValGluGly 198  
QY 1165 GCCAACCGTGAGATGACTTATTGT----- 1188  
Db 199 AlaPheTyrHisMetAsnTyrMetLeuGluAspGlySerArgValLeuGluValAlaThr 218  
QY 1189 -----AGCGACAGCTCTTCGGTGTGACCGAGGTGCTGTTTGCACAGTTGCCAAC 1242  
Db 219 ThrArgProGluThrMetPheGlyAspVal----- 228  
QY 1243 GCATTGGCTACAAATCCTGATGCTCATGCGGTATCCTCAAAATCGCTTGACGACAGTAC 1302  
Db 229 AlaValAlaValAsnProGlu-----AspProArgTyr 239  
QY 1303 AAAGACTACCTCCCTCG-----CTCGACCGTAAGGTGCTGCCGCCCATG 1347  
Db 240 LysAspLeuIleGlyLysAsnValIleLeuProIleAlaAsnLysLeuIleProIleVal 259  
QY 1348 CTCGAT-----ATTCTAGCCCGCGGT 1368  
Db 260 GlyAspGluHisAlaAspProGluPheGlyThrGlyValValLysIleThrProAlaHis 279  
QY 1369 ATCCCTGCGGACAAAGTCCCGCATATATTCAAGAATGTAAATCGACAAGAAATTCAAAGGC 1428  
Db 280 AspProAsnAspPheLeuValGlyGlnArgHisAsnLeuProGlnValAsnValMetAsn 299  
QY 1429 GACACGAGAAGTATGCAGACTTCGTATTTCGACAGAGTGTGGTTCCTTATAGCCAGCAG 1488  
Db 300 AspAspGlyThrMetAsnAspLeuAlaPheGluPheSerGlyMet-----AspArg 316  
QY 1489 TTCCATGCCATCTCAGTCCATGCGACAGGAAAGTTTGCCAAAGCTATCAGAAAGAT 1548  
Db 317 PheGluAlaArg-----LysAlaValAlaLysLeuGluGluIleGly 331  
QY 1549 CCGGCGAGTAGACTTTCCAAGACGCTAATAGTCTGCTCGCGCT----- 1593  
Db 332 AlaLeuValLysIleGluLysArgValHisSerValGlyHisSerGluArgThrGlyVal 351  
QY 1594 -----ATTCAGGCCGAT 1605  
Db 352 ValValGluProArgLeuSerThrGlnTrpPheValLysMetAspGlnLeuAlaLysAsn 371  
QY 1606 GCGATGGCCAAATGCCTATGCCATTCAGAGGCGCAGCGTCTTTCTTCCCGGTTTGGCT 1665  
Db 372 AlaIleAlaAsnGlnAspThrGluAspLys-----Val 382

QY 1666 GAGATGTACCCGCGAGCTGCTCTGCGCGAG-----GAT 1698-  
Db 393 GluPheTyrProProArgPheAsnAspThrPheLeuGlnTrpMetGluAsnValHisAsp 402  
QY 1699 GCCAACTTCACCATCGGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGT 1758  
Db 403 TrpValIleSerArgGlnLeuTrpTrpGlyHis-----GlnIlePro 416  
QY 1759 GCCTGGTACAACTATCATACAGCAAGGCGGTATTTGGAGAACCGAGGATCCTTAAGAC 1818  
Db 417 AlaTrpTyrAsn-----AlaAspGlyGluMetTyrValGlyGluGluAlaProGluGly 434  
QY 1819 GATGATTTTGGCGTACAGGAGAATATCTCTCGAC 1851  
Db 435 AspGlyTrpThrGlnAspGluAspValLeuAsp 445  
RESULT 3  
US-08-997-897-2  
; Sequence 2, Application US/08997897C  
; Patent No. 6114514  
; GENERAL INFORMATION:  
; APPLICANT: SRIVASTAVA, RANJANA  
; APPLICANT: KUMAR, DEEPAK  
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS SPECIFIC DNA FRAGMENT  
; FILE REFERENCE: u011469-7  
; CURRENT APPLICATION NUMBER: US/08/997,897C  
; CURRENT FILING DATE: 1997-12-24  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (4)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (6)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (20)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (29)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (54)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (64)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (69)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (89)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (99)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (114)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (119)  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (129)  
; FEATURE:  
; NAME/KEY: UNSURE





QY 739 TATTGCTCTGCTATTTCGGCCGCGGTTGTGGCAGCGGCATACACGCGGAATACGC 680  
|||  
Db 276 --CysArgArgCysPheGlyCysArgGlyCysArg----- 286  
QY 679 TGAAGTCGCGGTGTGACGGCGGCACATCCAGTTGTCGGTATCCGCCGCCAAGTTTACCTA 620  
|||  
Db 287 -----CysGlnGlyCys\*\*\*SerHisArgArgPheArgCysArgGluCysCys 302  
QY 619 CAGAGCTGGGAGGAGAAATACCATACGAACGCTCTTGAATACATCCTAGACGATGAGGA 560  
302 ----- 302  
QY 559 AGTATTGCTTTGTTGAATAGAAAGGCTCTACGATGCAGAGTTGTTCTCGTCTGCATTTT 500  
::: |||||  
Db 303 AsnCysArgCysTrp-----ArgCysArgGluCysSerArggProGly 317  
QY 499 CTT----- 497  
Db 318 LeuProGlyArgAsp\*\*\*ArgProValGlyHisArgLysIleProThrCysCysPheArg 337  
QY 496 -----TTTGGCCAGTTCTTGGCATACCTCTGAGCTTTCGAGCTTTCGCGCAGAC 455  
338 Cys\*\*\*ArgSerProArgSerArggProAlaLeu\*\*\*TrpProGlySerCys\*\*\*Thr 357  
QY 454 GCTCCATCTCGTCAGTATACCCCTTGAGCTGCTCTTCTACCTTTCGCTTACCTTTCACGA 395  
358 Asn-----ProIleArg 361  
QY 394 TCTTCCGCGAGATCTTACGGAAGACCCGGAATCGGAAGCTCTCACCCTATCGTCGCGAG 335  
362 CysCysProSer\*\*\*SerArgProIlePro-----AlaArgProArgLeuPro 377  
QY 334 AAACGAACCATCGCGCAGATAGTCGATCCACCGTCTTGGCTCTGGATACACCGT 275  
378 GlyArgSerArggTrp-----ProPro----- 385  
QY 274 ATCCGCGAGTGGTGT-----TGTAAAGATCAGGCCCTGATCGGACACTGTGA 227  
386 ThrLysSerGlyArgSerGlnAsnCysTrpHisArgSerSerGlySerArgThr----- 403  
QY 226 TACCGGTACATCCGCCACCGA-----AGATAACACCGGATTCGGCAATGGACGGCTGT 173  
404 ---ArgTyr---ArgHisArgCys\*\*\*ArgIleProThrThrAlaHisTrpSerSer\*\*\* 421  
QY 172 CGAAAC 167  
422 ArgSer 423

## RESULT 4

US-08-488-940-18  
Sequence 18, Application US/08488940  
Patent No. 5854049  
GENERAL INFORMATION:  
APPLICANT: Reed, Guy L.  
TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488, 940  
FILING DATE: 09-JUN-1995  
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Fraser, Janis K.  
REGISTRATION NUMBER: 34,819  
REFERENCE/DOCKET NUMBER: 05433/009001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-488-940-18  
Alignment Scores:  
Pred. No.: 0.00188 Length: 1194  
Score: 127.50 Matches: 133  
Percent Similarity: 35.05% Conservative: 99  
Best Local Similarity: 20.09% Mismatches: 263  
Query Match: 3.34% Indels: 167  
DB: 2 Gaps: 35

US-10-008-355-1 (1-2139) x US-08-488-940-18 (1-1194)

QY 223 GGTATCATCAGTCCGATCAGGCGCTG-----ATCTTTACCACAC 264  
Db 571 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 590  
QY 265 CACTGCGGATACGGTGTATCCAGAGCCAAAGCAGCGGTGATCAGCACTATCTGCGCGAT 324  
||| |||||  
Db 591 HisProGlyThrIleTyrGluArgAspSerSerIle-----ValThrHisAsp 607  
QY 325 GGTTCCTTTCTCCAGCATGGGTGAGGAGCTTCCGATCCG-----GGTCTTTCC 375  
||| |||||  
Db 608 AsnAspIlePheArgThrIle-----LeuProMetAspGlnGluPheThrTyrArg 624  
QY 376 GTGAAGTATCTCGCAGATCGTGAAGGTAAACGACAAAGTAGAAGCAGCAGCTCAAGGT 435  
|||||  
Db 625 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys-----SerGly 639  
QY 436 ATCACTGACGAGATGGAGCGTCTCGCAAGCTCAGGAGGTATGCCAAGAACATGCGCAAA 495  
::: |||||  
Db 640 LeuAsnGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 659  
QY 496 AAAGAA----- 501  
Db 660 GlyGluLysProTyrAspPropheAspArgSerHisLeuLysLeuPheThrIleLysTyr 679  
QY 502 ---AATGCAGACGAGAACCAACTCTGCATCTGATAGCGCTTCTATTCCAACACGAAATAC 558  
::: ||| |||||  
Db 680 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 699  
QY 559 -----TTCTCTATCATCGTATGATTCAGGACGCTTCGTATGGTATTTGCTCTCT 609  
||| |||||  
Db 700 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr----- 717  
QY 610 CCCAGCTCTAGTAAAGTTCGGA-----GCCGATACGGACAAC 648  
::: |||||  
Db 718 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 736  
QY 649 TGGATGTGGCGGCTCACACGGCGACTTCAGCGCTATTCCCGGTGATGCGGTGCCGAC 708  
|||  
Db 737 -----AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly----- 750  
QY 709 AACCGCGCGCGCAATACAGCAAGGACAATAAACCCCTTAAGCCCGTTTACTTCGCTGCC 768  
|||||  
Db 751 LysArgPro-----GluGlyGluAsnAlaSerTyrHis----- 761  
QY 769 GTATCCATGCAAGGCTACAGGCTGACGATATGCCATGACCATCGGTTTCCCGGCGAGT 828  
||| ||||| |||

Db 762 -----LeuAlaTyrAspAlaAspArgTyrThrGluGluArgGluValTyrSer 778  
QY 829 ACGATCGTACCTCACTTCTGGGTGTCGAGATCGTATCGAAACAGACAACAACTCT 888  
Db 779 TyrLeuArgTyr---ThrGlyThrProIleProAspAsnProAsnAspLysAsnAsnSer 797  
QY 889 CGTATC-----GAAGTTCCGGGTATCAAGCAAGGCATCTCGAAGGAA 930  
Db 798 GluLeuValValSerValAlaGlyThrValGluGlyThrAsnGlnAspIleSerLeuLys 817  
QY 931 GCCATGAGCGCATGAGGTACCGGTATCAATATGCCAGCAAGATGCTCAG----- 984  
Db 818 PhePheGluLeuLeuAspLeuThrSerArgProAlaHisGlyGlyLysThrGluGlnGlyLeu 837  
QY 985 -----AGTCTACTATTGGAAGANTTCGATCGGTATGAACCCGCTCTCGCTCGT 1035  
Db 838 SerProLysSerLysProPheAlaThrAspSerGlyAlaMetSerHisLysLeuGluLys 857  
QY 1036 CTTGACGTGATAGTCTGTAAGCGTCCGAGAGCAAGACATTCGAGACTGGATCCGTAAG 1095  
Db 858 AlaAspLeuLeu-----LysAlaIleGlnGluGlnLeuIleAlaAsn---ValHisSer 874  
QY 1096 AAC-----GGCAAGAGTGTCTCTATGCGGATGTTG 1128  
Db 875 AsnAspAspTyrPheGluValIleAspPheAlaSerAspAlaThrIleThrAspArgAsn 894  
QY 1129 TCTTCTCTGAAAGGCTTAAAGAGAGGCAAGCCAGCCGTAGATGACTTATTG 1188  
Db 895 GlyLysValTyrPheAlaAspLysAspGlySer---ValThrLeuProThrGlnProVal 913  
QY 1189 AGGAGAGCTCTCGGTGTGACGGAGGTGCTGTTTTCACAGTTTGCACAGCATG 1248  
Db 914 GlnGluPheLeuLeuSerGlyHisValArgValArgTyrLysGlu----- 928  
QY 1249 GCTACAATCTGATGCTCATGCTCGGTATCTCAAAATCGCTTGACGACAAGTAC----- 1302  
Db 929 -----LysProIleGlnAsnGlnAla-----LysSerValAspValGluTyrThrVal 944  
QY 1303 -----AAAGACTACTCTCCCTCGCTC-----GACCGTAAGTG 1335  
Db 945 GlnPheThrProLeuAsnProAspAspAspPheArgProGlyLeuLysAspThrLysLeu 964  
QY 1336 CTGCCCCCATGCTCGATATGTACGGCGGTATCCCTCCGACAGCTCCCGCATATA 1395  
Db 965 LeuLysThrLeu-----AlaIleGlyAspThrIleThrSerGlnGluLeuLeuAlaGln 982  
QY 1396 TTCAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGATGTCAGACTTCGTA 1455  
Db 983 AlaGlnSerIleLeuAsnLysAsnHisProGly-----TyrThrIleTyrGlu 998  
QY 1456 TTCGACAGAGTGTGTTCTTATAGCACAAGTTCATGCCATGCTCAAGTCCATGCAC 1515  
Db 999 ArgAspSerSerIleValThrHisAspAsnAspIlePheArgThrIleLeuProMetAsp 1018  
QY 1516 AAGAAAGATTGCCAAGGTATCGAAGAAGATCCGGCAGTAGACTTTCCAGAGCGTA 1575  
Db 1019 GlnGluPheThrTyrArgValLysAsnArgGluGlnAlaTyrArgIleAsnLysLysSer 1038  
QY 1576 ATAGTCTGCTCGCGTATTACGGCGCATGCGATGCCAATGCC---TATGCCATTGAG 1632  
Db 1039 GlyLeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLys 1058  
QY 1633 AAGGGCAAG-----GCTCTTTTCTTTGCGCGT 1659  
Db 1059 LysGlyGluLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLys 1078  
QY 1660 TTGCGTCAGATGTACCCCGGACGTGCTCGCGAGCATGCCAACTTCACC----- 1710  
Db 1079 TyrValAspValAspThrAsnGlnLeuLeuLysSerGluGlnLeuLeuThrAlaSerGlu 1098  
QY 1711 ATGGGTATGAGTACGGCTCCATCAAGGGATATACACCGGAGGCGGTGCC-----TGG 1764  
Db 1099 ArgAsnLeuAspPheArgAspLeu-----TyrAspProArgAspLysAlaLysLeuLeu 1116

QY 1765 TACAAC-----TATCATACGACGACGACGACGCGTATTG 1797  
Db 1117 TyrAsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLys----- 1133  
QY 1798 GAGAACGAGGATCCTAAGAGCGGATGAGTTTCCCGTACAGGAGATATATCCTCGACCTCTTC 1857  
Db 1134 -----ValGluAspAsnHisAspAspThrAsn 1142  
QY 1858 CGC-----ACCAAAAACTATGTCGTCG-----TATCCGAGAAACGCTCAGCTCCATATC 1905  
Db 1143 ArgIleIleThrValTyrMetGlyLysArgProGluGlyGluAsnAlaSerTyrHisLeu 1162  
QY 1906 GCTTTC 1911  
Db 1163 AlaTyr 1164  
RESULT 5  
US-09-059-584-53  
; Sequence 53, Application US/09059584  
; Patent No. 6440701  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patencin Release #1.0, Version #1.30  
; APPLICATION NUMBER: US/09/059,584  
; FILING DATE: 14-APR-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/778,570  
; FILING DATE: 03-JAN-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-794  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 713 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-059-584-53  
Alignment Scores:  
Pred. No.: 0.00313 Length: 713  
Score: 123.50 Matches: 123  
Percent Similarity: 31.34% Conservative: 76  
Best Local Similarity: 19.37% Mismatches: 269  
Query Match: 3.23% Indels: 167  
DB: 4 Gaps: 24

```
US-10-008-355-1 (1-2139) x US-09-059-584-53 (1-713)
QY 355 CTTCGGATTCCGGGTCTTTCCGTGAAGTATCTGGCAAGATCGTGAAGTAACGGACAAG 414
Db 156 IleProPheAspLysAsnLeuIleGluTyrLeuLysSerGluValValSerLys 175
QY 415 GTAAAGACACAGCTCAAGGGTATACACTGACGAGATGAGCGCTCGCCAAAGCTCAGGAG 474
Db 176 PheGluAlaGlnLysGlyGlyIleGluAsnThrArgLeuThrHisLysAspLeuSer 195
QY 475 GTATGCCAAGACTGGCCAAAGAAATGACAGACGACGAGAACCTCTGCATCGTAGAG 534
Db 196 SerGluGlnLysGluAlaLysValLysGluAlaLeuAspAsnAlaLeu----- 211
QY 535 CCTTCTATTCCAAACGAATACTCTCATCGTC-----TACGATGTATTCAAG 585
Db 212 ThrGlnPheAlaGlnGluLysTyrLysGluLeuIleGluAsnAlaHisAspLysLysSer 231
QY 586 GACGTTTCGTATGTPATTGTCTCCCTCCAGCTCTGTAGGTAAGTTTCGGAGCGATACGGAC 645
Db 232 AspAlaArg----- 234
QY 646 AACTGGATGTGGCCCGCTCACACGGCGCACTTCAGCGTATTCGCCGTGTATGCCGGTGCC 705
Db 234 ----- 234
QY 706 GACAAACGGCGCGCGGAATACAGCAAGGACAATAAACCTTATAAGCCGTTACTTCGCT 765
Db 235 ---AsnArgAspLeuGluTyrValLysSerGlyPheAsnTyr----- 247
QY 766 GCCGTATCCATGCAAGCTACAGGCTGACGACTATGCCATGACCATCGGTTTC----- 819
Db 248 -----LeuSerGlyTyrThrAlaThrAspHisAspLysLysThrAsnTyrArgGly 264
QY 820 -----CGGGCAGTAGC 831
Db 265 TyrTyrGlyAlaLeuTyrTyrLysGlySerGluThrAlaLysGluLeuProGlnThrSer 284
QY 832 GATCGCTACCTCACTCTCTGGGGT-----GTGGAAGATCGTATGCAAAACGACAACAA 885
Db 285 AlaLysTyrLysGlyTyrTrpAspPheMetThrAspAlaThrLeuAspAsnLysTyrThr 304
QY 886 CCTCGTATCGAAGTTCGGGTATC-----AAGCAAGCATCTGGAAGGAGCATGAGCGCA 942
Db 305 -----AspLeuProGlyIleAlaArgGlnThrGlnTrpArgSerLeuValSerThr 321
QY 943 GATCAG---GCTACCCGTATC-----AAATATGCCAGCAAG 975
Db 322 AspGluTyrAlaThrLeuLeuThrAspLysAsnAsnLysProSerAspTyrAsnGlyAla 341
QY 976 TATGCTCAGAGTGTCTAATATTGCAAGAAATTCGATCGGT----- 1014
Db 342 TyrGlyHisSerSerGluPheAspValAsnPheAlaAspLysLysIleLysGlyLysLeu 361
QY 1015 ATGAACCCGGTCTCGCTGCTTTGACGTGATAGTGTGCTAAGCGTCCGAGGAAAGCA 1074
Db 362 IleSerAsnGlnLeuSerGlyThrAlaValThrAlaLysGluArgTyrLysIleGluAla 381
QY 1075 TTCGACAGTGGATCCGTAAAGACGGCAAGAGTCTCTATGCGGATGATTGCTCTCT 1134
Db 382 AspIleHisGlyAsnArgPheArgGlySerAla-----ThrAla 394
QY 1135 CTCGAAAGGCTTATAAGGAAGGACCAAGGCCAACCCGTGAGATGACTTAT---TTGAGC 1191
Db 395 SerAspLysAla-----GluAspSerLysThrGlnHisProPheThrSerAspAlaThr 412
QY 1192 GAGACGCTCTTCGTGGTACGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1251
Db 413 AsnLysLeuGluGlyGlyPheTyrGlyProLysGlyGluGluLeuAlaGlyLysPheLeu 432
QY 1252 ACAATCTGTATGCTATCGCGGTATCCTCAATCGCTTGCAGCAAGTACAAAGACTAC 1311
Db 1311 ----- 1311
```

## RESULT 6

```
US-08-488-940-4
; Sequence 4, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
```

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA: US/08/488,940

APPLICATION NUMBER: US/08/488,940

FILING DATE: 09-JUN-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 05433/009001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 800 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-488-940-4

Alignment Scores:

Pred. No.: 0.00432 Length: 800  
 Score: 122.50 Matches: 130  
 Percent Similarity: 33.98% Conservative: 99  
 Best Local Similarity: 19.29% Mismatches: 254  
 Query Match: 3.21% Indels: 191  
 DB: 2 Gaps: 35

US-10-008-355-1 (1-2139) x US-08-488-940-4 (1-800)

QY 223 GGTATCACAGTCCGATCAGGGCTG-----ATCTTTACCAACAC 264  
 DB 177 GlyAspThrIleThrSerGlnGluLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 196  
 QY 265 CACTGCGGATACGGTGTATCCAGAGCAAGCAAGCGTGGATCAGACTATCTCGGGAT 324  
 DB 197 HisProGlyTyrThrIleTyrGluArgAspSerSerIle-----ValThrHisAsp 213  
 QY 325 GGTTCCTTCCTCCAGCATGGGTGAGGAGCTCCGATCCG-----GGTCTTTCC 375  
 DB 214 AsnAspIlePheArgThrIle-----LeuProMetAspGlnGluPheThrTyrArg 230  
 QY 376 GTGAGTATCTGCCAGATCGTGAAGTACGAGCAAGGTAGAGGACAGCTCAAGGT 435  
 DB 231 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys-----SerGly 245  
 QY 436 ATCACTGACGAGATGGAGCGTCCGCAAGCTCAGGAGGTATGCCAAGAACTGCCAAA 495  
 DB 246 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 265  
 QY 496 AAAGAA----- 501  
 DB 266 GlyGluLysProTyrAspPropPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 285  
 QY 502 ---AATGACAGAGAACCACTCGCATCGTAGAGCCTTCTATTCCAAACAGCAATAC 558  
 DB 286 ValAspValAspThrAsnGluLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 305  
 QY 559 -----TTCCCTCACTGCTACGATGATATTCAAGGAGCTTCGTATGTTGCTCCT 609  
 DB 306 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuTyr----- 323  
 QY 610 CCCAGCTCTGTAGGTAGTTCGGA-----GGCGATACGACAAC 648  
 DB 324 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 342  
 QY 649 TGGATGTGGCCGCTCACAGGGGCACTTCAGCGGTATTCGCGGTATGCGCGGTGCGAC 708

DB 343 -----AsnHisAspThrAsnArgIleIleThrValTyrMetGly----- 356  
 QY 709 AACCGCGCGCGAATACAGCAAGCAATAAACCCCTATAGCCCGTTACTTCGCTGCC 768  
 DB 357 LysArgPro-----GluGlyGluAsnAlaSerTyrHisLeuAlaTyr----- 370  
 QY 769 GTATCCATGCAAGGCTACAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGCGAGT 828  
 DB 371 -----Asp 371  
 QY 829 ACGATCGCTAC-----CTCAGCTCTCTGG 852  
 DB 372 LysAspArgTyrThrGluGluArgGluValTyrSerTyrLeuArgTyrThrGlyThr 391  
 QY 853 GGTGTGAAGATCGTATCGAAACGAGAACCAATCTCGTATC----- 894  
 DB 392 ProIleProAspAsnProAsnAspLysAsnSerGlnLeuValValSerValAlaGly 411  
 QY 895 GAAGTTCGCGGTATCAAGCAAGGATCTGGAAGGAGCAATGAGCGAGATCAGGCTACC 954  
 DB 412 ThrValGluGlyThrAsnGlnAspIleSerLeuLysPhePheGluIleAspLeuThrSer 431  
 QY 955 CGTATCAATATGCCAGCAATATGCTCAG-----AGTGCTAACTATTGG 999  
 DB 432 ArgProAlaHisGlyGlyLysThrGluGlnGlyLeuSerProLysSerLysProPheAla 451  
 QY 1000 AAGATTCGATCGATCAACCGGCTCGCTCTGTTGAGCTGATAGGTCGTAAAGCGT 1059  
 DB 452 ThrAspSerGlyAlaMetSerHisLysLeuGluLysAlaAspLeu-----LysAla 469  
 QY 1060 GCCGAGGAAGAGCATTCGACAGACTGGATCGTAAGAAC----- 1098  
 DB 470 IleGlnGluGlnLeuIleAlaAsn---ValHisSerAsnAspTyrPheGluValIle 488  
 QY 1099 -----GGCAAGATGCTGTCTATGGCGATGATTTGTCTTCGAAAAGGCTTATAAG 1152  
 DB 489 AspPheAlaSerAspAlaThrIleThrAspArgAsnGlyLysValTyrPheAlaAspLys 508  
 QY 1153 GAAGGAGCAAGGCGCAACCGGTGAGTACTTATTTGACGAGAGCGCTCTTCGGTGGTACC 1212  
 DB 509 AspGlySer---ValThrLeuProThrGlnProValGlnGluPheLeuLeuSerGlyHis 527  
 QY 1213 GAGTGTGTTCTTTGACAGTGTTCGCAACGATTTGGCTACAAATCCTGATGCTCATGCC 1272  
 DB 528 ValArgValArgTyrLysGlu-----LysProIleGlnAsnGln 540  
 QY 1273 GGTATCCTCAATCGCTTGACGACAAGTAC----- 1302  
 DB 541 Ala-----LysSerValAspValGluTyrThrValGlnPheThrProLeuAsnProAsp 558  
 QY 1303 AAGACTACCTCCCTCGCTC---GACGTAAGGTGCTGCCGCCATGCTCGATATTCTA 1359  
 DB 559 AspAspPheArgProGlyLeuLysAspThrLysLeuLeuLysThrLeu-----AlaIle 576  
 QY 1360 CGCCGGGTATCCCTCGCAGCAAGCTCCCGATATATTCAAGAATGTAAATCACAAGAAA 1419  
 DB 577 GlyAspThrIleThrSerGlnGluLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 596  
 QY 1420 TTCAAGCGGACAGAGAGTATGACAGCTTCGTATTCGACAGAGTGTGTTCTTAT 1479  
 DB 597 HisProGly-----TyrThrIleTyrGluArgAspSerSerIleValThrHis 612  
 QY 1480 AGCGAAGTTCCATGCCATGCTCAAGTCCATGGACAGGAAAGTTTGCAAGGCTATC 1539  
 DB 613 AspAsnAspIlePheArgThrIleLeuProMetAspGlnGluPheThrTyrArgValLys 632  
 QY 1540 GAGAAAGATCCGCGATAGAGCTTTTCCAAAGAGCGCTAATAGCTGCTCGCTGCTATTGAG 1599  
 DB 633 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 652  
 QY 1600 GCCGATGCGGATGGCAATGCC---TATGCCATTGAGAGGCGCAAG----- 1641

Db 653 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyLysProTyrAspPro 672  
Qy 1642 -----CGCTTTTTCGCGGTTTTCGCTGAGATGATACCCCGGAGGT 1683  
Db 673 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 692  
Qy 1684 GCTCTGCCGAGCGATGCCAAGTTCAC-----ATGCGTATGAGTACGGCTCCATC 1734  
Db 693 LeuLysSerGluGlnLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 712  
Qy 1735 AAGGATATGAAACCGCAGGCGGTCC-----TGGTACAAC----- 1770  
Db 713 -----TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 730  
Qy 1771 -----TATCATACACAGGCAAGGCGCTATTGGAGAAGCAGGATCCTAAGAGCGAT 1821  
Db 731 IleMetAspTyrThrLeuThrGlyLys----- 739  
Qy 1822 GAGTTTGGCGTACAGGAATATCTCGACCTCTTCCGC-----ACCAAAACTATGGT 1875  
Db 740 -----ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 756  
Qy 1876 CGC-----TATGCCGAGAACGGTCAGCTCCATATCGCTTC 1911  
Db 757 LysArgProGluGlyGluAsnAlaSerTyrHisLeuAlaTyr 770

RESULT 7  
US-08-488-940-3  
; Sequence 3, Application US/08488940  
; Patent No. 5854049  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Guy L.  
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/488,940  
; APPLICATION NUMBER: US/08/488,940  
; FILING DATE: 09-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraser, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 05433/009001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 813 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-488-940-3  
Alignment Scores:  
Pred. No.: 0.00438 Length: 813  
Score: 122.50 Matches: 130  
Percent Similarity: 33.98% Conservative: 99  
Best Local Similarity: 19.29% Mismatches: 254  
Query Match: 3.21% Indels: 191

DB: 2 Gaps: 35  
US-10-008-355-1 (1-2139) x US-08-488-940-3 (1-813)  
Qy 223 GGTATCACAGTCTCGATCAGGCGCTG-----ATCTTTTACCACACAC 264  
Db 190 GlyAspThrIleThrSerGlnGluLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 209  
Qy 265 CACTCGCGATACGGTGTATCCACAGCAACAGCGGTGGGATCAGACTATCTCGCGAT 324  
Db 210 HisProGlyTyrThrIleTyrGluArgAspSerSerIle-----ValThrHisAsp 226  
Qy 325 GGTTCCTGTTCTCGACGATGGTGGAGCTCCGATTCG-----GGTCTTTTCC 375  
Db 227 AsnAspIlePheArgThrIle-----LeuProMetAspGlnGluPheThrTyrArg 243  
Qy 376 GTGAAGTATCTGCCAGATCGTGAAGTAAACGACAGGTAGAGGACAGCTCAAGGT 435  
Db 244 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys-----SerGly 258  
Qy 436 ATCACTGACGAGATGGAGCGCTCGCAAGCTCAGGAGGTATGCCAAGAACTGCCAAA 495  
Db 259 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLys 278  
Qy 496 AAGAA----- 501  
Db 279 GlyLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 298  
Qy 502 ---AATGCAGACGAGAACCACTCTGCATCTGATGCTAGAGCTTTCTATTCACACAGAAATAC 558  
Db 299 ValAspValAspThrAsnGluLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 318  
Qy 559 -----TTCTCTCATCGTACGATCTATTCAAGGACGTCGTCTGTTATTTCTCTCT 609  
Db 319 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr----- 336  
Qy 610 CCCAGCTCTGTAGTAAAGTTCGGA-----GGCGATACGACGACAAC 648  
Db 337 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 355  
Qy 649 TGGATGTGGCGGCTCACACGGCGACTTCAGCGTATTTCCCGGTATGCGCGTGGCGAC 708  
Db 356 -----AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly----- 369  
Qy 709 AACCGCGCGCGAATACAGCAAGCAATAAACCTATAGCCGTTTACTTCTCGCTGCC 768  
Db 370 LysArgPro-----GluGlyGluAsnAlaSerTyrHisLeuAlaTyr----- 383  
Qy 769 GTATCCATGCAAGGCTACAGGCTCAGGACTATGCCATGACCATCGGTTTCCCGGCGAGT 828  
Db 384 -----Asp 384  
Qy 829 ACGATCGCTAC-----CTCAGCTCTTTGG 852  
Db 385 LysAspArgTyrThrGluGluGluArgGluValTyrSerTyrLeuArgTyrThrGlyThr 404  
Qy 853 GGTCTGGAAGATCGTATCGAAACAGAGCAACATCTCGTATC----- 894  
Db 405 ProIleProAspAsnProAsnAspLysAsnSerGlnLeuValValSerValAlaGly 424  
Qy 895 GAAGTTCCGGTATCAAGCAAGGATCTGGAAGGAGGACCATGAGCGGATCAGGCTACC 954  
Db 425 ThrValGluGlyThrAsnGlnAspIleSerLeuLysPheGluIleAspLeuThrSer 444  
Qy 955 CGTATCAATATGCCAGCATGCTCAG-----AGTGTCACTATTGG 999  
Db 445 ArgProAlaHisGlyLysThrGluGlnGlyLeuSerProLysSerLysProPheAla 464  
Qy 1000 AAGATTTCGATCGTATCAACCGCGGCTCTCGCTCTCTGACGTAGAGTCGTAAGCGT 1059  
Db 465 ThrAspSerGlyAlaMetSerHisLysLeuGluLysAlaAspLeu-----LysAla 482  
Qy 1060 GCCGAGGAAAGACATTTCGACACTGCGTAAAGAAC----- 1098

```
Db 483 ILeGlnGluGlnLeuIleAlaAsn---ValHisSerAsnAspAspTyrPheGluValIle 501
Qy 1099 -----GGCAAGAGCTGCTCTATGGCGATCTATTGCTCTCGAAAAAGGCTTATAAG 1152
Db 502 AspPheAlaSerAspAlaThrIleThrAspArgAsnGlyLysValTyrPheAlaAspLys 521
Qy 1153 GAAGGAGCCAGGCCAACCGTGAGATGACTTATTTGAGCGAGAGCGCTCTTCGGTGGTACC 1212
Db 522 AspGlySer---ValThrLeuProThrGlnProValGlnGluPheLeuLeuSerGlyHis 540
Qy 1213 GAGTGCTGCTTTTGACAGCTTGGCAACCGATGCTACAACTCTGATGCTCATGCC 1272
Db 541 ValArgValArgTyrLysGlu-----LysProIleGlnAsnGln 553
Qy 1273 GGTATCTCAAACTCGCTGACGACAAGTAC----- 1302
Db 554 Ala-----LysSerValAspValGluTyrThrValGlnPheThrProLeuAsnProAsp 571
Qy 1303 AAGACTACTCTCCCTCGCTC---GACCGTAAGTGCTGCCGCCATGCTCGATATTGTA 1359
Db 572 AspAspPheArgProGlyLeuLysAspThrLysLeuLeuLysThrLeu-----AlaIle 589
Qy 1360 CGCGCGGTATCCCTCGCGCAAGCTCCCGATATATCAAGATGTAAATCGACAAGAAA 1419
Db 590 GlyAspThrIleThrSerGlnGluLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 609
Qy 1420 TTCAAAGCGCACACGAAGATGATGACAGTTCGTATTTCGACAGAGTGTGGTCTCTAT 1479
Db 610 HisProGly-----TyrIleTyrGluArgAspSerSerIleValThrHis 625
Qy 1480 AGCGACAAAGTTCATGCTGCTCAAGTCCATGACAAAGAAAGTTTGCAAGGCTATC 1539
Db 626 AspAsnAspIlePheArgThrIleLeuProMetAspGlnGluPheThrTyrArgValLys 645
Qy 1540 GAGAAAGATCCGCGAGTAGAGTCTTCAAGAGCGTAAATAGTGTGCTCGCGCTATTACG 1599
Db 646 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 665
Qy 1600 GCCGATGCGATGGCCAATGCC---TATGCCATTGAGAGGCAAG----- 1641
Db 666 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyGluLysProTyrAspPro 685
Qy 1642 -----CGTCTTTTCTTCCCGGTTGCTGAGATGCTACCGCGGAGCT 1683
Db 686 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 705
Qy 1684 GCTCTGCGAGCGATGCCAATCTCAC-----ATGCGTATGACGTACGCGCTCCATC 1734
Db 706 LeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 725
Qy 1735 AAGGGATATGAACCGCAGGACGGTGC-----TGGTACAAC----- 1770
Db 726 -----TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 743
Qy 1771 -----TATCATACAGACAGGCAAGGCGCTATTTGGAGAAGCAGGATCTTAAGAGCGAT 1821
Db 744 IleMetAspTyrThrLeuThrGlyLys----- 752
Qy 1822 GAGTTGCGGTACAGGAGAATCTCTGACCTCTTCGC-----ACCAAAACTATGCT 1875
Db 753 -----ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 769
Qy 1876 CGC-----TATGCCGGAAGCGGTACGCTCCATATCGCTTTC 1911
Db 770 LysArgProGluGlyGluAsnAlaSerTyrHisLeuAlaTyr 783
```

## RESULT 8

```
us-08-488-940-2
; Sequence 2, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
```

```
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1181 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-488-940-2

Alignment Scores:
Pred. No.: 0.00589 Length: 1181
Score: 122.50 Matches: 130
Percent Similarity: 33.98% Conservative: 99
Best Local Similarity: 19.29% Mismatches: 254
Query Match: 3.21% Indels: 191
DB: 2 Gaps: 35

US-10-008-355-1 (1-2139) x US-08-488-940-2 (1-1181)
Qy 223 GGTATCATCAGTGTCCGATCAGGCGCTG-----ATCTTTACCAACAC 264
Db 558 GlyAspThrIleThrSerGlnGluLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 577
Qy 265 CACTGCGGATACGGTGTCTATCCAGAGCCAAAGCAGCGGTGGATCAGCACTATCTGCGCGAT 324
Db 578 HisProGlyTyrThrIleTyrGluArgAspSerSerIle-----ValThrHisAsp 594
Qy 325 GGTTCGTTTCTCCGACCATGGTGTGAGGAGCTTCGATTCG-----GGTCTTTCC 375
Db 595 AsnAspIlePheArgThrIle-----LeuProMetAspGlnGluPheThrTyrArg 611
Qy 376 GTGAAGTATCTCGCAAGATCGTGAAGTAAAGGACGAGTAGAAGGACAGCTCAAGGT 435
Db 612 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys-----SerGly 626
Qy 436 ATCACTGACGAGATGAGCGCTGCGCAAGAGCTCAGGAGGTATGCCAAGAACTGCCCAA 495
Db 627 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 646
Qy 496 AAGAA----- 501
Db 647 GlyGluLysProTyrAspPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 666
Qy 502 ---AATGCAGACGAGAACCAACTCTGCATCTGATGCTAGAGCCTTTCTATTCCAACAACGAATAC 558
Db 667 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 686
```





; MOLECULE TYPE: protein  
US-08-488-940-1

## Alignment Scores:

Pred. No.: 0.00594 Length: 1194  
Score: 122.50 Matches: 130  
Percent Similarity: 33.98% Conservative: 99  
Best Local Similarity: 19.29% Mismatches: 254  
Query Match: 3.21% Indels: 191  
DB: 2 Gaps: 35

US-10-008-355-1 (1-2139) x US-08-488-940-1 (1-1194)

QY 223 GGTATCACAGTGTCCGATCAGGCGCTG-----ATCTTTACCAACAC 264  
DQ 571 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 590  
QY 265 CACTGCGGATACGGTCTATCCAGAGCCCAAGCAGCGTGATCAGCACTATCTGCGCGAT 324  
DQ 591 HisProGlyTyrThrIleTyrGluArgAspSerSerIle-----ValThrHisAsp 607  
QY 325 GGTTCGTCTTCGCGACGATGGGTGAGGAGCTTCCGATTCG-----GGTCTTTCC 375  
DQ 608 AsnAspIlePheArgThrIle-----LeuPrometAspGlnGluPheThrTyrArg 624  
QY 376 GTGAAGTATCTCGCAAGATCGTGAAGGTAAACGAGCAAGGTAGAGCAGCACTCAAGGGT 435  
DQ 625 ValLysAsnArgGluGlnAlaTyrArgIleAsnLysLys-----SerGly 639  
QY 436 ATCACTGACGAGATGAGCGCTGCGCAAGCTCAGGAGGTATGCCAAGNACTGGCCAAA 495  
DQ 640 LeuAsnGluGluIleAsnAsnThrAspLeuIleSerGluLysTyrTyrValLeuLysLys 659  
QY 496 AAGAA-----501  
DQ 660 GlyLysProTyrAspProPheAspArgSerHisLeuLysLeuPheThrIleLysTyr 679  
QY 502 ---AATGACAGAGAACCAACTCTGCATCGTAGAGCCTTCTATTCCAAACAGCAATAC 558  
DQ 680 ValAspValAspThrAsnGluLeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArg 699  
QY 559 -----TTCCTCATCTCTACGATGATTTCAAGGACGCTTCGTATGGTATTTGCTCCT 609  
DQ 700 AsnLeuAspPheArgAspLeuTyrAspProArgAspLysAlaLysLeuLeuTyr----- 717  
QY 610 CCCAGCTCTCTAGGTAGTTCGGA-----GGCGATACGGACAAC 648  
DQ 718 ---AsnAsnLeuAspAlaPheGlyIleMetAspTyrThrLeuThrGlyLysValGluAsp 736  
QY 649 TGGATGTGCGCGCTCAGCGGCGACTTCAGCGTATTCGCGTGTATTCGCGTGGCGGAC 708  
DQ 737 -----AsnHisAspAspThrAsnArgIleIleThrValTyrMetGly----- 750  
QY 709 AACCGCGCGCCGAATACACAGCAGGACAATAACCTATTAAGCCGTTTACTTCGCTGCC 768  
DQ 751 LysArgPro-----GluGlyGluAsnAlaSerTyrHisLeuAlaTyr----- 764  
QY 769 GTATCATGCAAGGCTACAAGGCTGAGGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828  
DQ 765 -----Asp 765  
QY 829 ACGGATCGCTAC-----CTCACTTCTTGG 852  
DQ 766 LysAspArgTyrThrGluGluArgGluValTyrSerTyrLeuArgTyrThrGlyThr 785  
QY 853 GGTGTGAACATCGTATCGAAACAGGAGACAATCTTCGTATC-----894  
DQ 786 ProIleProAspAsnProAsnAspLysAsnAsnSerGlnLeuValValSerValaGly 805  
QY 895 GAATTCGCGGTATCAAGCAGGCTCTGGAAGGAGGCCATGATGAGCGCAGATCAGGCTACC 954  
DQ 806 ThrValGluGlyThrAsnGlnAspIleSerLeuLysPhePheGluIleAspLeuThrSer 925

QY 955 CGTATCAAAATATGCCAGCAAGTATGCTCAG-----AGTGCTACTACTATTGG 999  
DQ 826 ArgProAlaHisGlyGlyThrGluGlnGlyLeuSerProLysSerLysProPheAla 845  
QY 1000 AAGAATTCGATCGGTATGAACCGCGCTCGCTCGCTCTTACGCTGATAGTCTGAAGCGT 1059  
DQ 846 ThrAspSerGlyAlaMetSerHisLysLeuGluLysAlaAspLeu-----LysAla 863  
QY 1060 GCGAGGAAAGAGCATTCGACAGCTGATCCGCTAAGAAC-----1098  
DQ 864 IleGlnGluGlnLeuIleAlaAsn---ValHisSerAsnAspAspTyrPheGluValIle 882  
QY 1099 -----GGCAAGAGTGTCTATGCGGATGATTGCTTCTCTCGAAAAGGCTTATAAG 1152  
DQ 883 AspPheAlaSerAspAlaThrIleThrAspArgAsnGlyLysValTyrPheAlaAspLys 902  
QY 1153 GAAGGAGCAAGCCCAACCTGAGATCACTTATTGAGCGAGACGCTCTTCGCTGTGCTACC 1212  
DQ 903 AspGlySer---ValThrLeuProThrGlnProValGlnGluPheLeuLeuSerGlyHis 921  
QY 1213 GAGGTGTTGTTTGCACAGCTTGGCCACGCACTGGCTACAAATCCTGATGCTCATGCC 1272  
DQ 922 ValArgValArgTyrLysGlu-----LysProIleGlnAsnGln 934  
QY 1273 GGTATCCTCAAAATCGCTTGACGACAGTAC-----1302  
DQ 935 Ala-----LysSerValAspValGluTyrThrValGlnPheThrProLeuAsnProAsp 952  
QY 1303 AAGACTACTCTCCCTCGCTC---GACCGTAAGTGTGCTGCGCCCATGCTCGATATTGTA 1359  
DQ 953 AspAspPheArgProGlyLeuLysAspThrLysLeuLeuLysThrLeu-----AlaIle 970  
QY 1360 CGCGCGGTATCCTCGCGACAAGCTCCCGATATATTCAAGATGTAAATCGACAAGAAA 1419  
DQ 971 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 990  
QY 1420 TTCAAGGCGACACGAGAAGTATGCACTTCGTATTTCGACAGAGTGTGGTTCTTAT 1479  
DQ 991 HisProGly-----TyrThrIleTyrGluArgAspSerSerIleValThrHis 1006  
QY 1480 AGCGACAAGTTCCATGCCATGCCATCAAGTCCATCGACAGGAGAAAAGTTTGCAGGCTATC 1539  
DQ 1007 AspAsnAspIlePheArgThrIleLeuProMetAspGlnGluPheThrTyrArgValLys 1026  
QY 1540 GAGAAAGATCCGCGAGTAGAGCTTCCAGAGCGTAATAGCTGCTCGCTGCTATTTCAG 1599  
DQ 1027 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 1046  
QY 1600 GCGATGCGGATGCCAATGCC---TATGCCATTGAGAGGAGCAAG-----1641  
DQ 1047 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyGluLysProTyrAspPro 1066  
QY 1642 -----CGCTTTTCTTTCGCGTTTGGCTGAGTGTACCCCGGACGT 1683  
DQ 1067 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 1086  
QY 1684 GCTCTCGCGAGCGATGCCAATTCAC---ATGCGTATGAGTACGGCTCCATC 1734  
DQ 1087 LeuLeuLysSerGluGlnLeuLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 1106  
QY 1735 AAGGATATGAACCGCAGGACGCTGCC-----TGGTACAAAC-----1770  
DQ 1107 -----TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 1124  
QY 1771 -----TATCATCAGCAGCAGGCGGTATTTGGAGAGCAGGATCTTAAGAGCGAT 1821  
DQ 1125 IleMetAspTyrThrLeuThrGlyLys-----1133  
QY 1822 GAGTTTCCCTACAGGAGATATCTCGACCTCTTCGCG-----ACCAAAACATATGGT 1875  
DQ 1134 -----ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 1150  
QY 1876 CGC-----TATGCCGAGAACGGTCACTCATATCATCTTTC 1911



```
QY 1360 CGCGGGGTHATCCCTGCCGACAGCTCCCGCATATATTCAGAAATGTAATCGACAAGAAA 1419
Db 971 GlyAspThrIleThrSerGlnGluLeuLeuAlaGlnAlaGlnSerIleLeuAsnLysAsn 990
QY 1420 TTCAAAGGCGCACAGCAAGAGTATGACAGACTTCGTATTCGACAGAGTGGTTCCTTAT 1479
Db 991 HisProGly-----TyrThrIleTyrGluArgAspSerIleValThrHis 1006
QY 1480 AGGCACAAGTCCATGCCATGCTCAAGTCCATGACAGCAAGAAAGTTTGCAAGGCTATC 1539
Db 1007 AspAsnAspIlePheArgThrIleLeuLeuProMetAspGlnGluPheThrTyrArgValLys 1026
QY 1540 GACAAGATCCGCGAGTACAGCTTCCAGAGCGGTATACAGTCTGCTCGCGGTATTCAG 1599
Db 1027 AsnArgGluGlnAlaTyrArgIleAsnLysLysSerGlyLeuAsnGluGluIleAsnAsn 1046
QY 1600 GCCGATGCGATGCCAATGCC---TATGCCATTGAGAAGGGCAAG----- 1641
Db 1047 ThrAspLeuIleSerGluLysTyrTyrValLeuLysLysGlyGluLysProTyrAspPro 1066
QY 1642 -----CCTCTTTCTTTTGGCGGTTTGGCGTGAGATGTACCCCGGACGT 1683
Db 1067 PheAspArgSerHisLeuLysLeuPheThrIleLysTyrValAspValAspThrAsnGlu 1086
QY 1684 GCTCTCGCGAGCGATGCCAATCTCAC-----ATGCGTATGAGCTACGGCTCCATC 1734
Db 1087 LeuLysSerGluGlnLeuLeuThrAlaSerGluArgAsnLeuAspPheArgAspLeu 1106
QY 1735 AAGGGATATCAACCGGAGGAGCGGTGC-----TGGTACAAAC----- 1770
Db 1107 -----TyrAspProArgAspLysAlaLysLeuLeuTyrAsnAsnLeuAspAlaPheGly 1124
QY 1771 -----TATCATACGACAGCAGGCGGTATTTGGAGAAGCAGGATCCTTAAGAGCGAT 1821
Db 1125 IleMetAspTyrThrLeuThrGlyLys----- 1133
QY 1822 GAGTTTCCGTCAGGAGATATCTCGACCTCTTCGCG-----ACCAAAACTATGGT 1875
Db 1134 -----ValGluAspAsnHisAspAspThrAsnArgIleIleThrValTyrMetGly 1150
QY 1876 CGC-----TATGCCGAGAAGCGGTACGCTCCATATCGCTTTC 1911
Db 1151 LysArgProGluGlyGluAsnAlaSerTyrHisLeuAlaTyr 1164

RESULT 11
US-09-036-987A-3
; Sequence 3, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R.
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2152 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-036-987A-3

Alignment Scores:
Pred. No.: 0.0213 Length: 2152
Score: 119.00 Matches: 161
Percent Similarity: 30.96% Conservative: 100
Best Local Similarity: 19.10% Mismatches: 274
Query Match: 3.01% Indels: 308
DB: 4 Gaps: 33

US-10-008-355-1 (1-2139) x US-09-036-987A-3 (1-2152)
QY 2139 TTAGATCAACTTCAGCTCTTGGATGACAGCGGGGCGACTGACCCCATTTGTCAATCATCAA 2080
Db 1415 LeuAspGlyLeuAlaLeuValAspGluProThrAlaThrAlaProLeuGlyAspGlyGlu 1434
QY 2079 GAGAACGTAGCGGATGTCACGCT-----GATTGTGCGCTCGACATCGGG 2035
Db 1435 ValArgIleAlaMetArgAlaAlaGlyValAsnPheArgAspAlaLeuIleAlaLeuGly 1454
QY 2034 TTGCAACTCGATGTACCACTCATAGCTTCCAGTTCGCCATTCGCAAGCAAGACGCGTACG 1975
Db 1455 Met-----TyrProGlyValAlaSerLeuGlySerGluGly 1466
QY 1974 ACGGCCCTTCTTATCGAATAC-----GGGCTACCGGAGTTACCGCCCGTGTGTT 1921
Db 1467 AlaGlyValValValGluThrGlyProGlyValThrGlyLeuAlaProGlyAsp----- 1484
QY 1920 GTTCGATAGGAAGACGATATGGAGCTGACCGTCTCGGCATAGCACCATAGTTTGTGTT 1861
Db 1485 -----ArgValMetGlyMetIleProLysAlaPheGly 1495
QY 1860 CGGGAAGAGTCCGAGGATATTCCTCTGTACGGCAACTC----- 1822
Db 1496 ProLeuAlaValAlaAspHisArgMetValThrArgIleProAlaGlyTrpSerPheAla 1515
QY 1821 -----ATCGCTCTTAGGATCCTGCTTCTCCAATACGCCCTTGCCTGTCGTATG 1774
Db 1516 ArgAlaAlaSerValProIleValPheLeuThrAlaTyrTyrAlaLeuValAspLeuAla 1535
QY 1773 ATAGTTGTACAGGACCGCTCTCGGTTTCATCCCTTGTATGAGGCGGTAGCTCATACG 1714
Db 1536 GlyLeuArgProGlyGluSerLeuLeuValHisSerAlaAlaGlyGlyValGly----- 1553
QY 1713 CATGGTGAAGTTGGCATCGCTCGGCGAGACGACGTCCGGGGTACATCTACCCAAACCGCG 1654
Db 1554 MetAlaAlaIleGlnLeuAlaArgHisLeuGlyAlaGluValTyrAlaThrAlaSerGlu 1573
QY 1654 ----- 1654
Db 1574 AspLysTrpGlnAlaValGluLeuSerArgGluHisLeuAlaSerSerArgThrCysAsp 1593
QY 1653 ---AAAGAAAGACGCTTGCCCTTCTCAATGGCATAGGATGGCCATCGCATCGGCTG 1597
Db 1594 PheGluGlnGlnPheLeuGlyAlaThrGlyGlyArgGlyValAsp-----ValValLeu 1611
QY 1596 AATAGCGGAGCAGCAGCTATTACGCTCTTGGAAAGCTCTACTGCCGAGATCTTTCTCGAT 1537
```



; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 2152

; TYPE: PRT

; ORGANISM: Saccharopolyspora spinosa

US-09-370-700-3

## Alignment Scores:

Pred. No.:	0.0213	Length:	2152
Score:	119.00	Matches:	161
Percent Similarity:	30.96%	Conservative:	100
Best Local Similarity:	19.10%	Mismatches:	274
Query Match:	3.01%	Indels:	308
DB:	4	Gaps:	33

US-10-008-355-1 (1-2139) x US-09-370-700-3 (1-2152)

QY	2139	TTAGATCAACTTCAGCTCTGGATGACAGCGGGCCACTGACCCCATTTGTCATCATGAA	2080
Db	1415	LeuAspGlyLeuAlaLeuValAspGluProThrAlaThrAlaProLeuGlyAspGlyGlu	1434
QY	2079	GAGAACGTAGCGGATGCCACGCT-----GATTGTGGCTGCAGATCGGG	2035
Db	1435	ValArgIleAlaMetArgAlaAlaGlyValAsnPheArgAspAlaLeuIleAlaLeuGly	1454
QY	2034	TTGGAACGTGATGCACCACTCATAGCTTCCAGTTGCCATGCGAAAGCAAGACCGATCAG	1975
Db	1455	Met-----TyrProGlyValAlaSerLeuGlySerGluGly	1466
QY	1974	ACGCCGTTCTTATCGAATAC-----GGCGTACCGGAGTTACCGCCCGTGTGCTGT	1921
Db	1467	AlaGlyValValGluThrGlyProGlyValThrGlyLeuAlaProGlyAsp-----	1484
QY	1920	GTTGATAGGAAGCGATATGGAGCTTCCGCGTATCCGCGATAGCGACCATAGTTTGGT	1861
Db	1485	-----ArgValMetGlyMetIleProLysAlaPheGly	1495
QY	1860	GCGGAAGGTGCGAGGATATCTCTGTACGCGCAACTC-----	1822
Db	1496	ProLeuAlaValAlaAspHisArgMetValThrArgIleProAlaGlyTrpSerPheAla	1515
QY	1821	-----ATCGCTCTTAGATCCTCTCTCCAAATACGCCCTTGCTGCTGCTATG	1774
Db	1516	ArgAlaAlaSerValProIleValPheLeuThrAlaTyrTyrAlaLeuValAlaAspLeuAla	1535
QY	1773	ATAGTTTACACGACCGCTCGGTTTCATATCCCTTGATGAGCGGTAGCTCATACG	1714
Db	1536	GlyLeuArgProGlyGluSerLeuLeuValHisSerAlaAlaGlyGlyValGly-----	1553
QY	1713	CATGGTGAAGTTGGCATCGCTCGCGACGACGCTCGCGGTGTACATCTCACGCAACCGCG	1654
Db	1554	MetAlaAlaIleGlnLeuAlaArgHisLeuGlyAlaGluValTyrAlaThrAlaSerGlu	1573
QY	1654	-----	1654
Db	1574	AspLysTrpGlnAlaValGluLeuSerArgGluHisLeuAlaSerSerArgThrCysAsp	1593
QY	1653	---AAGAAAGAGCGCTTGCCTCTCAATGGCATAGCATGTCATCGCATCGGCGCTG	1597
Db	1594	PheGluGlnPheLeuGlyAlaThrGlyArgGlyValAsp-----ValValLeu	1611
QY	1596	AATAGCGGACGACGACGTATTCGCTCTTGAAAGCTCTACTCCCGGATCTTCTCGAT	1537
Db	1612	AsnSerLeuAlaGlyGlu-----PheAlaAsp	1620
QY	1536	AGC-----CTTGGCAAACTTTTCTCTGTCCATGGA-----	1507
Db	1621	AlaSerLeuArgMetLeuProArgGlyArgPheLeuGluLeuGlyLysThrAspVal	1640
QY	1507	-----	1507
Db	1641	ArgAspProValGluValAlaAspAlaHisProGlyValSerTyrGlnAlaPheAspThr	1660

QY	1506	CTTTCAGCATGCGCATGGAACTTGTGCTATTAAGGAACACACACTCTTGTGCA-----TACGAA	1450
Db	1661	ValGluAlaGlyProGlnArgIleGlyGluMetLeuHisGluLeuValGlnLeuPheGlu	1680
QY	1449	GTCTGCATACCTT-----CTTCGTGTCGCTTTGAATTT	1417
Db	1681	GlyArgValLeuGluProLeuProValThrAlaTrpAspValArgGlnAlaProGluAla	1700
QY	1416	CTTTCGATTACATT---CTTGAATATATCGGAGAGTGTCTTTCGTCTGTCGTCACGCGATT	1360
Db	1701	LeuArgHisLeuSerGlnAlaArgHisValGlyLysLeuValLeuThrMetProVal	1720
QY	1359	TACAAT-----ATCGACATGCGCGG-----	1339
Db	1721	TrpAspAlaAlaGlyThrValLeuValThrGlyGlyThrGlyAlaLeuGlyAlaGluVal	1740
QY	1338	---CAGCACCTTACGCTCGAGCGAGGGAGGTAGTCTTTTCTACTGTGTCGTCACGCGATT	1282
Db	1741	AlaArgHisLeuValIleGluArgGlyValArgAsnLeuValLeuValSerArgGly	1760
QY	1281	GAGGATACCGCATGAGCATCAGGATTTGTAGCCAATCGCTTGGC-----	1237
Db	1761	ProAlaAlaSerGlyAlaAlaGluLeuValAlaGlnLeuThrAlaTyrGlyAlaGluVal	1780
QY	1236	-----AACTGTGCAAAACGACCACTCGGTACCAACGAGAGCTGCTGCTCAATA	1183
Db	1781	SerLeuGlnAlaCysAspVal-----AlaAspArgGluThrLeuAlaLysVal	1796
QY	1182	AGTCATCTCAGGTTGGCTTGGCTCTCTTATAGCTTTTCGAGAGAGAGCAATAC	1123
Db	1797	-----LeuAlaSerIleProAspGluHisProLeu	1806
QY	1122	ATGCCATAGACAGCACTTTCGCGTTCTTACGATCGAGTCTCGGAATGCTCTTCTC	1063
Db	1807	ThrAlaValValHisAlaAlaGlyValLeuAspGlyValSerGlu-----	1822
QY	1062	GGCAGCTTACGACCTATCAGTC-----AAG	1036
Db	1823	---SerLeuThrValGluArgLeuAspGlnValLeuArgProLysValAspGlyAlaArg	1841
QY	1035	ACGAGCCAGACCGCGTTCATACCGATCGAATCTTCCATAGTTAGCACTCTCAGGATA	976
Db	1842	AsnLeuLeuGluLeuIleAspProAspValAlaLeuValLeuPheSerSerValSerGly	1861
QY	975	CTTCTGCGCATATTTGATAGCGGTAGCTGCTCGCTCATGCTTCTTCCAGATGCC	916
Db	1862	ValLeuGlySerGlyGlyGlnGlyAsnTyrAlaAlaAlaAsnSerPheLeu---AspAla	1880
QY	915	TTGCTT-----GATACCGCGAACTTCGATACGAGGATGTTCTC-----	877
Db	1881	LeuAlaGlnGlnArgGlnSerArgGlyLeuProThrArgSerLeuAlaTrpGlyProTrp	1900
QY	877	-----	877
Db	1901	AlaGluHisGlyMetAlaSerThrLeuArgGluAlaGlnAspArgLeuAlaArgSer	1920
QY	876	-----GTTTTCGATACGATCTTCCACACC	853
Db	1921	GlyLeuLeuProIleSerThrGluGluGlyLeuSerGlnPheAspAlaAlaCysGlyGly	1940
QY	852	CCAAGAGTAGGTAGCG-----ATCCGTACTGCCCGGGAACCGATGTGTATGGCATAGTC	796
Db	1941	AlaHisThrValValAlaProValArgPheSerArgLeuSerAspGlyAsnAlaIleLys	1960
QY	795	GTCAGCTTGTAGCTTGCATGGATACGGCAGCGAAGTAACAGCGCTTATAGGCTTATT	736
Db	1961	PheSerVal-----LeuGlnGlyLeuValGly-----	1969
QY	735	GTCCTTGTGTATTTCGCGCGCGGTTGTCGCGCACCGGCATACACGCGGAATACGCTGAA	676
Db	1969	-----	1969
QY	675	GTCGCCGCTGTGACGCGGCCACATCCAGTTGTCGCTATGCTCCGAACTTACCTACAGA	616





Qy 1291 ----- 1291  
Db 507 ysgluasnproglyserilegtyrGlnValGlnAlaAsnSerCysSerGlyHisGlyT 527  
Qy 1292 -----ACGA 1295  
Db 527 hrSerMetCysGlyileCysAsnCysAspSerTyrPheGlyAsnLysCysGluCysS 547  
Qy 1296 CAAGTACAAGACTACTCC-----CCTCGCT 1322  
Db 547 erAlaThrAspLeuThrSerLysPheAlaAsnAspThrSerCysArgAlaAspSerThrS 567  
Qy 1323 CGACCGTAAGGTGTCGCCCGCATCGCATATTGTAGCGCGCGTATCCCTGCCGACAA 1382  
Db 567 erThrThrAspCys-SerGlyArgGlyHisCysCysValGlyAlaCysGluCysHisLys 586  
Qy 1383 GCTCCCCATATATCAA-----GAATGTAATCGACAAGAAATT 1421  
Db 587 ArgProasnProIleGluIleSerGlyLysHisCysGluCysAspAsnPheSerCys 606  
Qy 1422 CAAGGGCGACACGAAGAGTATG-----AGA 1448  
Db 607 GluArgAsnArgAsnGlnLeuCysSerGlyProAspHisGlyThrCysGluCysGlyArg 626  
Qy 1449 CTTCGTATTCGA-----CAAGAGTGTGTTCTTATAGCGACAAGTTCCA 1493  
Db 627 CysLysCysLysProGlyTrpThrGlySerAsnCysGlyCysGlnGluSerAsnAspThr 646  
Qy 1494 TGCCATGCTCAAGTCCATCGACAGGAAGAAAGTTGCCAAGCTATCGAAGAAAGATCCGGC 1553  
Db 647 Cys-----MetProGlyGlyGlyGluIleCysSerGlyHis----- 659  
Qy 1554 AGTAGAGCTTTCCAAGACGTAATAGTGTCTCGCGCTATTTCAGGCGCATGCGAT--- 1610  
Db 660 -----GlyThrCysGluCys 664  
Qy 1611 GGCCAAATGCTATGCCATGAGAGGCAAGCGTCTTTTCTTTTCCCGCTTTGCGTGAGAT 1670  
Db 665 GlyValCysLysCysThrValAsnAspGlnGlyArgPheSerGlyArgHisCysGluLys 684  
Qy 1671 GTACCCCGAGCTGCTGCGCGAG-----CGATGCCAACTT 1706  
Db 685 CysPro---ThrCysSerSerGlyArgCysGlnGluLeuLysAspCysValGlnCysGlnMet 703  
Qy 1707 CACCATGCGTATGAGTACGCGCTCCATCAAGGATATGAACCGCAGGACGCTGC----- 1760  
Db 704 TyrLysThrGlyGluLeuLys-----AsnGlyAspAspCysAlaArgAsnCysThrGln 721  
Qy 1761 CTGGTACAACATCATACGACGCAAGGCGTATTGAGGAAGCAGGATCCTTAAGACGGA 1820  
Db 722 PheValProValGlyValGluLys-ValGluIleAspGluThrLysAspGluGlnMetCys 741  
Qy 1821 TGAGTTTCCCGTACAGGAGATATCTCTCGACTCTTCCGCGCACCACAAACTATGTCGCTA 1880  
Db 741 sLysPhePheAspGluAspAspCysLysPheMetPhe-----LysTyr 755  
Qy 1881 TGCCGAGAACGGTACGCTCCATATCGCTTTCTCTATCGAACACAGAC----- 1926  
Db 755 rSerGluGlnGlyGluLeuHisValTyrAlaGlnGluAsnLysGluCysProAlaLysVa 775  
Qy 1927 -----ATCAGCGCGCTAACTCCGGTATCGCGTATTCGATAAAGACGGCGG 1973  
Db 775 lPheMetLeuGlyIleValMetGlyValIleAlaAlaIleVal----- 789  
Qy 1974 TGTGATCGCTTCTGCTTTCGATGCGCACTGGGAGCTATGAGT 2016  
Db 790 -LeuValGlyLeuAlaIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 803  
RESULT 14  
US-08-938-085A-33  
; Sequence 33, Application US/08938085A  
; Patent No. 6339148

## GENERAL INFORMATION:

APPLICANT: Sheppard, Dean  
APPLICANT: Quatanta, Vito  
APPLICANT: Pytela, Robert  
TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses  
TITLE OF INVENTION: Thereof  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938.085A  
FILING DATE: 26-SEP-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728,215  
FILING DATE: 11-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Parent, Annette S.  
REGISTRATION NUMBER: 42,058  
REFERENCE/DOCKET NUMBER: 023070-080210US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 846 amino acids  
TYPE: amino acid  
STRANDEDNESS: linear  
TOPOLOGY: protein  
MOLECULE TYPE: protein  
US-08-938-085A-33

## Alignment Scores:

Pred. No.: 0.0227 Length: 346  
Score: 115.50 Matches: 119  
Percent Similarity: 30.99% Conservative: 56  
Best Local Similarity: 19.93% Mismatches: 201  
Query Match: 3.02% Indels: 211  
DB: 4 Gaps: 26

US-10-008-355-1 (1-2139) x US-08-938-085A-33 (1-846)

Qy 642 GGACAACTGGATGTGGCGCGTCACACGGCGACTTCAGCGTATTCGCGTGTATGCCGG 701  
Db 282 GlyAsnLeuAsp-----AlaProGluGlyGlyPheAspAlaIleMetGlnAlaIle 299  
Qy 702 TGCCGCAACACCGCGCGCGAATACAGCAAGACATAAACCCCTAAGCCCGCTTACTT 761  
Db 300 CysArgSerGlnIleGlyTrpArgGluGlnAlaArg-----ArgLeuLeu 314  
Qy 762 CGCTGCCGTATCCATGTCAAGGC-----TACAAAGCTGACGACTATGCCATGACCATCG 815  
Db 315 ValPheSerThrAspAla-GlyPheHisTyrAlaGlyLysLeuGlyValIrl 334  
Qy 816 TTTCCCGCGCAGTACCGATCGCTACCTCCTCTTGGGGTGTGGAGATCGTATCGAAAA 875  
Db 334 eAlaProAsnAspGlyGluCysHisLeuSerProLysGlyGluTyrThrHisSerThr 354  
Qy 876 CGAGAACAAATCCTCGTATCGAAGTTTCGCGGTATCAAGCAAGGCATCTGAAGGAAGCCAT 935  
Db 354 uGlnAspTyrProSerIleSer-----GlnIleAsnGlnLysValLysAspAsnAlaIrl 372  
Qy 936 GAGCCGAGATCAGGCTACCCGCTATCAAAATATGCCAGCAAGTATGCTCAGAGTGTAACTA 995



MOLECULE TYPE: protein  
US-08-484-105-14

## Alignment Scores:

Pred. No.: 0.0264 Length: 885  
Score: 115.00 Matches: 158  
Percent Similarity: 30.61% Conservative: 108  
Best Local Similarity: 18.18% Mismatches: 316  
Query Match: 3.01% Indels: 287  
DB: 1 Gaps: 34

US-10-008-355-1 (1-2139) x US-08-484-105-14 (1-885)

```

Qy 90 CTTCAACGAACCTCAATCAGGAGAATCTGGATCGAATGGTGGCTGAGCTCGGCTTTTACGCTCCC 149
Db 28 ProArgArgArgArgArgGlyAspAlaThrGluHisGluValIleAsnLeuValArgTyr 47
Qy 150 GTTGATTGCTCTACAGTTCGACAGCGTCCATTCGCAATGCCAATGCCGTTATCTTCGG 209
Db 48 AspGlyValArgLeuTyrProGlyValThrIleValCysLysValGluGlyAlaAspGlu 67
Qy 210 TGGCGGATGTACCGTATCAGATCTCCGATCAGGCGCTGATCTTTACCAACCACCTG 269
Db 68 LeuSerAlaTyrMetIleHisGluValArgLeuAsnThrSerAsnTyr-----Val 84
Qy 270 CGGATACGGTGC----- 281
Db 85 GluLeuTyrCysLeuAsnTyrLeuSerTrpTyrGluIleAsnAlaAlaGluArgTyrLys 104
Qy 282 -----TATCAGAGCCAAAGCAGCGTGATCAGCTACTCTCGCGGA 323
Db 105 GluLeuAspGlyGluPheTyrThrAsnLysGluLysGlyAspLys-PhePheGluG1 124
Qy 324 TGGTTTCCTTCTCCAGATGGGTGAGAGCTTCCTCCGATTCGGGT---CTTTCGTGAA 380
Db 124 uThrPheAlaSerGlnSerIleLysAsnGluLeuTyrLeuThrAlaGluLeuSerGluI1 144
Qy 381 GTATCTGCGCAAGATCGTGAAGGTAAACGACAGAGTAGAAGGACAGCTCAAGGTTATCAC 440
Db 144 eTyrLeuArgAspLeuGlnPheValAlaAsn-----IleLys 156
Qy 441 TCACGAGATGGAGCTCTGCCAAAGCTCAGGAG-----GT 476
Db 156 sAsnGluLysGluTyrLeuAspSerValAsnGluGlyLysMetAspSerAsnMetPheLe 176
Qy 477 ATGCCAAGAACTGGCCAAAAGAAATGCAGACAGAGCAACCAACTCTGCATC----- 528
Db 176 uCysArgSerAlaCysLeuProSerGlyThrAsnLeuAlaAspLeuAspIleHisPhePh 196
Qy 529 -GTAGAGCCTTCTATTCCAAACAAGTAATCTCTCATCTGTACGATGTATTCAAGGA 587
Db 196 eGluGluLysIleArgSerSerAsnProLysValSerLeuGluTyr-----LeuArgAs 214
Qy 588 CGTTCGTATGTATTGCTCTCTCCAGCTCTGTAGTAGTTCGGAGGCGATACGGACAA 647
Db 214 pIleThrLeu----- 217
Qy 648 CTGGATGTGGCGGCTCACAGGGGAGCTTCAGCGTATTCCGCGTATCCGCGTCCGGA 707
Db 218 -----ProLysLeuProLysProLeuAsnLysSerLysValHisAlaArgGluLys 734
Qy 708 CAACCGCGCGCGGAATACAGCAGGACAAATAAACCCCTATAAGCCCGTTTACTTCGCTGC 767
Db 234 sValValAlaThrLysLeuGlnSerAspAsnThrProSerLysLysSerPheGlnGlnTh 254
Qy 768 CGTATCCATGCAGGCTACAGGCTGACGATATCCCATGACCATC----- 813
Db 254 rValSerLysThrAsnAlaGluValGlnArgIleAlaSerThrIleValAsnGluLysG1 274
Qy 814 -----GGTTTCCCGGCGAGTACGGATCGCTACCT 842
Db 274 uAlaIleSerAspAsnGluSerAspLeuSerGluTyrHisGluSerLysGluGluPheAl 294

```

```

Qy 843 CACTTCTTGGGTGTGGAAGATCGTATCGAAACGAGAACAACTCTCTGATCGAA----- 897
Db 294 aAsnAlaSerSerSerAspSerAspGluGluPheGluAspTyrGlnSerAlaGluGluLe 314
Qy 898 -----GTTCCGCGGTATCAAGCAAGGCAATCTGGAAGGA 929
Db 314 uAlaIleValGluProAlaLysLysLysValArgSerIleLysProAspIle-----Pr 332
Qy 930 AGCCATGAGCCAGATCAGGCTACCCGTATCAAAATATCCAGCAAGTAGTGTCTCAGAGTGC 989
Db 332 oIleSerProValLysSerGlnThrProLeuGlnProSerAlaValHisSerSerProAr 352
Qy 990 TAACTATTGGAAGATTCGATCGGTATGAACCGCGTCTCGTCTGTGACGTGATAGG 1049
Db 352 gLysPhePheLysAsnAsnIle----- 359
Qy 1050 TCGTAAGCGTCCGAGAGAAAGAGCAATTCGCACACTGGATCCGT-----AAGAACGGGAA 1103
Db 360 ----ValArgAla---LysLysAlaTyrThrProPheSerLysArgTyrLysAsnProLy 377
Qy 1104 GAGTCTCTCTATGCGCATGTA-----TTCTCTTCTCT 1136
Db 377 sIleProAspLeuAsnAspIlePheGlnArgHisAsnAsnAspLeuAspIleAlaAlaLe 397
Qy 1137 CGAAAGCGCTTATAAGGAAGGAGGCAAGCCAAACCGTGAGATGACTTATTTGAGCGGAC 1196
Db 397 uGluGluArgPheArgThr---ValSerAlaLysGlyLysMet-----GluTh 412
Qy 1197 GCTTCTCTGCTGTGATACC-----GAGTGGTTCG 1223
Db 412 rIlePheSerLysValLysLysGlnLeuAsnSerArgAsnSerLysGluGluIleValLy 432
Qy 1224 TTTTTCAGAGTTTGCACAGCATTCGCTACAAATCCTCATGCTCATCGCGGTATCCTCAA 1283
Db 432 sAlaAlaAspPheAspAsnTyrLeuProAlaArgGluAsnGluPheAlaSerIleTyrLe 452
Qy 1284 ATCGCTT----- 1290
Db 452 uSerLeuTyrSerAlaIleGluAlaGlyThrSerThrSerIleTyrIleAlaGlyThrPr 472
Qy 1291 -----GACGACAGGTACAAAGACTACCTCCCTCCGCTCGA 1325
Db 472 oGlyValGlyLysThrLeuThrValArgGluValValLysAspLeuMetThrSerAlaAs 492
Qy 1326 CGGTAAAGTGTGTCGCGCCCATG-----CTCGATATTGTAGCGCG 1364
Db 492 pGlnLysGluLeuProArgPheGlnTyrIleGluIleAsnGlyLeuLysIleValLys--- 511
Qy 1365 GCGTATCCTCGCCGACAAAGCTCCCGCATATATTCAGAATGTAAATC-----GACAAGAA 1418
Db 512 -----AlaSerAspSerTyrGluValPheTrpGlnLysIleSerGlyGluLysLe 528
Qy 1419 ATTCAAAGCGGACACGACAGATATGCAGACTTCGTATTCGACAGAGTGTGGTTCCTTA 1478
Db 528 uThrSerGlyAlaAlaMetGluSerLeuGluPheTyrPheAsnLys-----ValPro--- 545
Qy 1479 TAGCGACAAGTTCCATGCCATCTCAAGTCCATGCACAGCAAGAAAGTTTGCCAAAGCTAT 1538
Db 546 -----AlaTh 547
Qy 1539 CGAGAAGATCCGGCAGTAGAGCTTTCCAAAGACGCGTAATAGTGTGCTGCGCGCTATTCA 1598
Db 547 rLysLysArgProIleValValLeuLeuAspGluLeuAspAlaLeuValSerLysSerG1 567
Qy 1599 GGCCGATCGGATGGCCAATGGCTAT-----GCCATTGAGAAGGGCAACGCTCTTTT 1649
Db 567 n---AspValMetTyrAsnPhePheAsnTrpAlaThrTyrSerAsnAlaLeuLeuIleVa 586
Qy 1650 CTTTTCGCT-----GGTTTGGTGGATGTACCCCGGACGCTGCTCTCCCGAG 1694
Db 586 lValAlaValAlaAsnThrLeuAspLeuProGluArgHisLeuGlyAsnLysIleSerSe 606
Qy 1695 CGATGCCAACTTCACC----- 1710

```

```

Db 606 rArgIleGlyPheThrArgIleMetPheThrGlyThrHisGluGluLeuArgThrIl 626
Qy 1711 -----ATGCGTATGAGCTACGCTCCATCAAGGATAT-----GAACCGCAGGACGG 1757
Db 626 eIleAsnLeuArgLeuLysTyrLeuAsnGluSerPheTyrValAspProGluThrGl 646
Qy 1758 TGCCTGTGTACAATATCATACGACAGGCAAGGGCTATTGGAGAGCAGCATCCTAAGAG 1817
Db 646 ySerSerTyrMetIleSerProAspSerSerThrIleGluThrAspGluGluLysAr 666
Qy 1818 CGATGAGTTCCGCTACAGGAGATATCCTC----- 1848
Db 666 gLysAspPheSerAsnTyrLysArgLeuLysLeuArgIleAsnProAspAlaIleGluIl 686
Qy 1849 -----GACCTCTTCGCG-----ACCAA 1865
Db 686 eAlaSerArgLysIleAlaSerValSerGlyAspValArgAlaLeuLysValValLy 706
Qy 1866 AAATATGCTGCTATGCCGAGACAGGTCAGCTC-----CATATCGCTTT 1910
Db 706 sArgAlaValGluTyrAlaGluAsnAspTyrLeuLysArgLeuArgTyrGluArgLeuVa 726
Qy 1911 CCTATCGAACACGACATCAGCGCGGTAACCTCCGT----- 1947
Db 726 lAsnSerLysLysAspThrSerGlyAsnGlyThrGlyAsnGluGluLeuGlnSerValGl 746
Qy 1948 -----AGCCCGCTATTTCGATAAGAACGGCCGCT 1976
Db 746 uIleLysHisIleThrLysAlaLeuAsnGluSerSerThrSerProGluGlnPheIl 766
Qy 1977 GATCGCTCTTGCTTCGATGGC----- 1998
Db 766 eSerGlyLeuSerPheSerGlyLysLeuPheLeuTyrAlaLeuIleAsnLeuIleLysLy 786
Qy 1999 -----AACTGGGAAGCTATGAGTGGTGACATCGAGTTCGACCCGATCTCGACCGCAAT 2054
Db 786 sLysGlnThrAspValGlnLeuGlyAspIleValGluGluMetArgLeuLeuIleAspVa 806
Qy 2055 CAGCGTGGACATCCGCTACGTTCTC 2079
Db 806 lAsnGlyAsnAsnLysTyrIleLeu 814

```

Search completed: May 16, 2003, 12:39:14  
Job time : 73.5 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:35:41 ; Search time 83.5 Seconds  
(without alignments)  
4941.515 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 3820

Sequence: 1 atgcaaatgaataaaag.....aagagctgaagtgtatctaa 2139

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 362588 seqs, 96450795 residues

Total number of hits satisfying chosen parameters: 725176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US10008355/runat\_16052003\_110341\_9460/app\_query.fasta\_1.2311  
-DB=PublishedApplications\_AA\_QFMT=fastan -SUFFIX=n2p.rapp -MINMATCH=0.1  
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10008355 -CGN\_1\_1\_25\_@runat\_16052003\_110341\_9460  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAPP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pcp:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pcp:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pcp:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pcp:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pcp:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pcp:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pcp:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pcp:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pcp:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pcp:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pcp:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3719	97.4	712	9	US-10-008-355-2
2	3652	95.6	699	9	US-10-008-355-8
3	1276	33.4	720	9	US-10-008-355-9
4	907.5	23.8	716	9	US-10-008-355-7

Alignment Scores:	7.63e-310	Length:	712
Pred. No.:	3719.00	Matches:	712
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.36%	Indels:	0

ALIGNMENTS

RESULT 1  
US-10-008-355-2  
; Sequence 2, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 712  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-2

5	884.5	23.2	734	9	US-10-008-355-5	Sequence 5, Appli
6	807.5	21.1	732	9	US-10-008-355-6	Sequence 6, Appli
7	274	7.2	52	9	US-10-008-355-3	Sequence 3, Appli
8	148.5	3.9	1528	9	US-09-945-917-3	Sequence 3, Appli
9	148.5	3.9	1583	9	US-09-945-917-4	Sequence 4, Appli
10	145	3.8	1367	10	US-09-801-368-108	Sequence 108, App
11	133	3.5	883	10	US-09-815-242-13382	Sequence 13382, A
12	127	3.3	883	10	US-09-815-242-13384	Sequence 13384, A
13	122.5	3.2	26	9	US-10-008-355-18	Sequence 18, Appli
14	118.5	3.1	2042	9	US-10-192-584-6	Sequence 6, Appli
15	118	3.1	1814	10	US-09-920-552-103	Sequence 103, App
16	117	3.1	708	10	US-09-925-301-1012	Sequence 1012, App
17	116.5	3.0	1169	10	US-09-801-368-106	Sequence 106, App
18	116	3.0	1138	12	US-10-036-328A-6	Sequence 6, Appli
19	116	3.0	1210	12	US-10-036-328A-2	Sequence 2, Appli
20	115.5	3.0	846	9	US-10-072-841-33	Sequence 33, Appli
21	115.5	2.9	1270	9	US-10-101-464A-979	Sequence 979, App
22	114	3.0	2701	9	US-10-171-311-83	Sequence 83, Appli
23	113	3.0	687	9	US-09-969-384-27	Sequence 27, Appli
24	112	2.9	1014	9	US-10-097-340-4	Sequence 4, Appli
25	111.5	2.9	682	10	US-09-920-552-13	Sequence 13, Appli
26	111.5	2.9	1023	9	US-09-884-696-5	Sequence 5, Appli
27	111	2.9	1114	10	US-09-782-874-2	Sequence 2, Appli
28	111	2.9	5877	9	US-10-142-515-11	Sequence 11, Appli
29	110	2.9	2039	9	US-10-192-584-7	Sequence 7, Appli
30	109.5	2.9	845	9	US-10-027-806-62	Sequence 62, Appli
31	109.5	2.9	845	9	US-10-034-623-62	Sequence 62, Appli
32	109.5	2.9	845	9	US-10-027-801-62	Sequence 62, Appli
33	109	2.9	630	9	US-10-213-990-45	Sequence 45, Appli
34	109	2.9	2382	9	US-10-196-935A-2	Sequence 2, Appli
35	107.5	2.7	890	9	US-10-101-464A-958	Sequence 958, App
36	107.5	2.8	1394	9	US-09-945-917-57	Sequence 57, Appli
37	107.5	2.8	1449	9	US-09-945-917-58	Sequence 58, Appli
38	107	2.8	889	9	US-09-909-567B-44	Sequence 44, Appli
39	107	2.8	1607	10	US-09-938-275-10	Sequence 10, Appli
40	106.5	2.8	523	10	US-09-815-242-11918	Sequence 11918, A
41	106	2.8	556	10	US-09-801-368-426	Sequence 426, App
42	106	2.8	972	9	US-10-100-049-23	Sequence 23, Appli
43	105.5	2.8	5179	9	US-10-025-380-1068	Sequence 1068, Ap
44	105.5	2.8	5179	10	US-09-922-217-1068	Sequence 1068, Ap
45	105.5	2.8	5179	10	US-09-833-263-1068	Sequence 1068, Ap

DB: 9 Gaps: 0  
US-10-008-355-1 (1-2139) x US-10-008-355-2 (1-712)  
QY 1 ATGCAATGAATTAATAAGATTTCTTCCTCGGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60  
DB 1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaLaLeuLeuLeuGlyAlaSerGly 20  
QY 61 GTAGCCAAAGCCGCAAAAGGCATGTGGCTCTCAACGAACATCAATCAGGAGCAATCTGGAT 120  
DB 21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlnLeuAsnGlnLeuAsn 40  
QY 121 CGAATCGCTGAGCTCGGCTTTACGCTCCGCTTGGATTGCTCTACAGTTTCACAAAGCCG 180  
DB 41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60  
QY 181 TCCATTGCCAATGCCGCGGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGAT 240  
DB 61 SerIleAlaAsnAlaValAlaPheGlyGlyCysThrGlyIleThrValSerAsp 80  
QY 241 CAGGCGCTGATCTTTACCAACCACCATGCGGATACGGTGTATCCAGAGCCAAAGCACG 300  
DB 81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100  
QY 301 GTGATACAGCACTATCTCGCGGATGGTTTCGTTCTCGCAGATGGGTGAGAGCTTCCG 360  
DB 101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuPro 120  
QY 361 ATTCGGGTCTTCCGTGAAGTATCTCGCAAGATCTGGAAGGTAAACGGACAAGGTAGAA 420  
DB 121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140  
QY 421 GGACAGCTCAAGGGTATCACATGACGAGATGAGCGTCTGCGCAAGCTCAGGAGTATGC 480  
DB 141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160  
QY 481 CAAGAAGTGGCCAAAAGAAAATGACAGACGAGAACCAACTCTGCATGCTAGAGCCTTTC 540  
DB 161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180  
QY 541 TATTCCAAACAGTAATCTTCCTCATCTCTACGATGATTTCAAGGAGCTTCGTATGGTA 600  
DB 181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200  
QY 601 TTTGCTCTCCAGCTCTGTAGTAAGTTTCGGAGCGCATACGGACAACATGGATGTGGCGG 660  
DB 201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220  
QY 661 CGTCACAGCGCGGACTTCAGCGTATTCGGCGTGTATCGCGGTGCGGACAAACCGCGCGCC 720  
DB 221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240  
QY 721 GAATACACAGGACAATAAACCTTATAAGCCGTTTACTTCGCTGCGGTATCCATCAAA 780  
DB 241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260  
QY 781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGCGCAGTACCGATCGGTAC 840  
DB 261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280  
QY 841 CTCACTTCTTGGGTGGGAGATCGTATCGAANAACGAGACAATCTCTCGTATCGAAGTT 900  
DB 281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300  
QY 901 CGCGGTATCAAGCAAGCATCTGGAAGAACCCATGAGCGCAGATCAGGCTACCGGTATC 960  
DB 301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320  
QY 961 AAATATGCCAGCAAGTATGCTCAGAGTGCCTAACTATTGGAGAATTCGATCGGTATGAAC 1020  
DB 321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340  
QY 1021 CGCGGTCTCGCTCTTGACGTATAGGTGTAAGGCTGCCGAGGAAGACAGCATTCGCA 1080

DB 341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360  
QY 1081 GACTGGATCCGTAAGACCGCAAGAGTGTCTCTATGCGGATGATTGTCTCTCTCTCGAA 1140  
DB 361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380  
QY 1141 AAGCTTTATAAGGAAGGAGCCAAAGCCGTCAGCTGAGCTGACTTTATTGAGGAGACGCTC 1200  
DB 381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400  
QY 1201 TTCGGTGGTACCGAGGTGGTTGCTTTGTCACAGTTGCCAACGCATTTGGCTACAAATCTCT 1260  
DB 401 PheGlyGlyThrGluValAlaArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420  
QY 1261 GATCCTCATGCCGCTATCCTCAAAATCGCTGACGACAGTACAAGACTACTCTCCCTCG 1320  
DB 421 AspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSer 440  
QY 1321 CTCGACCGTAAAGTCTGCCCGCATGCTCATATTGCTACGCCGCGCTATCTCCCTCCGAC 1380  
DB 441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp 460  
QY 1381 AAGCTCCCCGATATATTCAAGAATGTATTCGACAGAANAATTCAAAGGCGACACGAAGA 1440  
DB 461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480  
QY 1441 TATCAGACTTCGTATTCGCAAGAGTGTGTTCTTATAGGCACAAAGTTCATGCGCCATG 1500  
DB 481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500  
QY 1501 CTCGAAGTCCATGCAAGAAAGTTTGCAGAGCTATTCGAGAAGATCCGCGCATAGAG 1560  
DB 501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520  
QY 1561 CTTTCCAAAGACGGTAATAGTGTGCTGCGCGCTATTCAGGCGCGATGGCATGCCAATGCC 1620  
DB 521 LeuSerLysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540  
QY 1621 TATCCCATTTGAGAAGGCAAGCGCTTTCTTCTCCGCTTTCGCTGAGATCTACCCCGGA 1680  
DB 541 TyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrProGly 560  
QY 1681 CGTCTCTCGCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740  
DB 561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580  
QY 1741 TATCAACCGCAGGACGGTGCCTGGTACAACTATCATACGACAGCAAGGCGGTATTGGAG 1800  
DB 581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrGlyLysGlyValLeuGlu 600  
QY 1801 AAGCAGGATCCTAAGAGCGATGAGTTTGCCTACAGGAGAATATCTCGACCTCTTCCGC 1860  
DB 601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620  
QY 1861 ACCAAAACATATGCTCGCTATGCCGAGAACGGTCAGCTCCATATCGCTTTCTCTATCGAAC 1920  
DB 621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640  
QY 1921 AACACATACCGCGCGGTAACTCCGCTAGCCCGGTATTCGATTAAGACGCGCGCTCATC 1980  
DB 641 AsnAspIleThrGlyLysSerGlySerProValPheAspLysAsnGlyArgLeuIle 660  
QY 1981 GGTCTTCTTTCGATGGAAGTATAGTGTGATGATGAGTGTGATGAGTGTGGAACCGCAT 2040  
DB 661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680  
QY 2041 CTCGAGCGCAAAATCAGCGTGGACATCCGCTACGCTTCTTTCATGATTGACAAATGGGGT 2100  
DB 681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700  
QY 2101 CAGTGCCTCCGCTCTCATCCAAAGAGCTCAAGTTGATC 2136  
|||||

Db 701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712

## RESULT 2

US-10-008-355-8  
; Sequence 8, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Babula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 699  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-8

Alignment Scores:  
Pred. No.: 4,19e-304 Length: 699  
Score: 3652.00 Matches: 699  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 95.60% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-10-008-355-8 (1-699)

QY	1	ATGCAATGAATTAATAAGTATCTTCGAGCAGCCCTGCTGGTGGTTCAGGG	60
Db	1	MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuGlyAlaSerGly	20
QY	61	GTAGCCAAAGCCGACAAAGGCATGTGGCTCTCAACGAATCAATCAGGAGAATCTGGAT	120
Db	21	ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlnLeuAsnLeuAsp	40
QY	121	CGAATGGTGAGTCGGGTTTACGCTCCGTTGGATTGCTCTACATTTGCGACAACCGG	180
Db	41	ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro	60
QY	181	TCCATTGCCAATGCGGTGTTATCTTCGGTGGCGGATGATCCGGTATCACAGTCTCCGAT	240
Db	61	SerIleAlaAsnAlaValIlePheGlyGlyCysThrGlyIleThrValSerAsp	80
QY	241	CAGGGCCTGATCTTTACCAACCACTGCGGATACGGTACGTATCCAGAGCCAAAGCACG	300
Db	81	GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr	100
QY	301	GTGATCACGACTATCGCGGATGTTTCTGCTCGCAGATGGGTGAGGACTTCGCG	360
Db	101	ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluLeuPro	120
QY	361	ATTCGGGTCTTCCGTGAAGTATCTCGCAAGATCGTGAAGGTAACGGACAAGGTAGAA	420
Db	121	IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu	140
QY	421	GGACAGCTCAAGGATACCTGACGAGATGGAGCGTCTGCCAAAGCTCAGGAGGTATGC	480
Db	141	GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys	160
QY	481	CAAGAAGTGGCCAAAGAAAATGCGACGAGAACCAACTCTCGATCGGTAGAGCCCTTC	540
Db	161	GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe	180
QY	541	TATTCACACACGAAATCTTCCTCATCTGCTACCATGATATTCAAGGAGCTTCGTATGGTA	600
Db	181	TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal	200

QY	601	TTTGCTCTCCAGCTCTGTAGTTAAGTTCCGAGCGGATACGGACAACCTGGATGGCCG	660
Db	201	PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro	220
QY	661	CGTCACACGGGCGACTTCACGGTATTCCGGGTGTATCCGGTCCGCGACACCGCCGCG	720
Db	221	ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla	240
QY	721	GAATACAGCAAGGACAAATAAACCTATAAGCCGTTTACTTCGCTCCGCTATCCATCAA	780
Db	241	GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln	260
QY	781	GGCTACAAAGCTGACGACTATGCGCATGACCATCGGTTTCCCGGCGACGATCGGTAC	840
Db	261	GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr	280
QY	841	CTCACTTCTTGGGTGTGGAGATCGTATCGAAGCAACGAGACAATCTCTGTATCGAAGTT	900
Db	281	LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluVal	300
QY	901	CGGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATC	960
Db	301	ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle	320
QY	961	AAATATGCCACAGATATGCTCAGAGTCTAACTATTGGAAGAAATTCGATCGGTATGAAC	1020
Db	321	LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn	340
QY	1021	CGGGTCTCGCTCTGTGAGCTGATAGGTGCTAAGCGTCCGAGGAAAGACGATTCGCA	1080
Db	341	ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArgAlaPheAla	360
QY	1081	GACTGGATCCGTAAAGACGCAAGAGTCTCTATGCGGATGATTGTCTTCTCTCAAA	1140
Db	361	AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu	380
QY	1141	AAGCTTATTAAGGAAGGAGCCCAAGCCGTCGATGAGTACTTATTTGAGCGAGACGCTC	1200
Db	381	LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu	400
QY	1201	TTCCGCTGCTACCGAGGTGGTTCGTTTTCACAGTTCGCCAAGCATGGCTACAAATCCT	1260
Db	401	PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro	420
QY	1261	GATCCTCATGCGGTATCCTCAAAATCGCTTCAACGACAGTACAAAGACTACTCCCTCG	1320
Db	421	AspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSer	440
QY	1321	CTCAGCGCTAAGGTGCTGCCCGCATGCTCGATATTGTACCGCGCGTATCCCTCCGAC	1380
Db	441	LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp	460
QY	1381	AAGCTCCCGCATATTCAGAATGTATCGACAAGAAATTCAAAGGGGACACAGAGAAG	1440
Db	461	LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys	480
QY	1441	TATCAGACTTCGTATTTCGACAAGAGTGTGTTCTCTATAGCGACAAGTTCCTCATGCCATG	1500
Db	481	TyrAlaAspPheValPheAspLysSerValProTyrSerAspLysPheHisAlaMet	500
QY	1501	CTCAAGTCCATGGCAAGGAAAAGTTTGGCAAGGCTATCGAGAAGATCCCGCAGTAGAG	1560
Db	501	LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu	520
QY	1561	CTTTCCAAAGAGCGTAATAGCTGCTCGCGCTTATTCAGCGCGGATCGATGCCAATGCC	1620
Db	521	LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla	540
QY	1621	TATCCCATTTGAAGAGGCAAGCGCTTTTCTTCTCCGCTTTGCGTGAGATGATCCCGGA	1680
Db	541	TyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrProGly	560



```
Qy 1681 CGTGCTCTCGCGAGCGATGCGCAACTTCACCATGCGTATGAGTACGGCTCCATCAAGGGA 1740
Db 561 ArgAlaLeuProSerAspAlaAsnPhetThrMetArgMetSerTyrGlySerIleLysGly 580
Qy 1741 TATGACCGCAGGCGGTGCGTGTACAACTATCATACGACGAGCGAGGGGTATTGGAG 1800
Db 581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600
Qy 1801 AAGCAGGATCTTAAGAGCGATGAGTTTCCGTACAGGAGATATCTCTCGACTCTTCGCG 1860
Db 601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620
Qy 1861 ACCAAAACTATGTCGCTATGCGGAGAACGGTCAAGCTCCATATCGCTTCTTCTATCGAAC 1920
Db 621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640
Qy 1921 AACGACATCAGCGGCGGTAACTCCGGTAGCCCGCTATTCGATAGAACGGCGCTCTGATC 1980
Db 641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660
Qy 1981 GGTCTTGTCTTTCGATGGCAACTGGGAGCTATGAGTGGTGACATCGAGTTTCAAACCCGAT 2040
Db 661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680
Qy 2041 CTGACGCGCACAACTCAGCGTGGACATCGCGTACGCTTCTCTTCATGATTGACAAATGG 2097
Db 681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrp 699

RESULT 3
US-10-008-355-9
; Sequence 9, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.0040101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
; US-10-008-355-9

Alignment Scores:
Pred. No.: 2,07e-100 Length: 720
Score: 1276.00 Matches: 283
Percent Similarity: 56.53% Conservative: 128
Best Local Similarity: 38.93% Mismatches: 286
Query Match: 33.40% Indels: 30
DB: 9 Gaps: 11

US-10-008-355-1 (1-2139) x US-10-008-355-9 (1-720)
Qy 16 AAAAGTATCTTCGAGGACGCCCTGCTGTGGTGGTTCACGGGTAGCCAAAGCCGAC 75
Db 3 LysArgLeuLeuLeuProLeuPheAlaValLeuLeuCysLeuCysGlnIleAlaHisAlaAsp 22
Qy 76 AAAGCATGTGGTCCCTCAAGCACTCAATCAGGAGATCTGGATCGAATCGGTGAGCTC 135
Db 23 GluGlyMetTrpLeuMetGlnGlnLeuGlyArgLys--TyrAlaGlnMetLysGluArg 41
Qy 136 GCGTTTACGCTCCCGTGGATTCTCTACAGTTTCGACAGCGCGTCCATTGCCAATGCC 195
Db 42 GlyLeuLysMetLysGluTyrAspLeuTyrAsnProAsnGlyThrSerLeuLysAspAla 61
Qy 196 GTGGTTATCTTCGTGGGGATGTACCGGTATCATCAGTGTCCGATCAGGGCCTGATCTTT 255
```

```
Db 62 ValValLeuPheAspGlyGlyCysThrGlyGluValValSerAspArgGlyLeuValLeu 81
Qy 256 ACCAACCACTCGGATCGGTGCTATCCAGACCAACGCGGTGCGTACGAGCTAT 315
Db 82 ThrAsnHisLysCysGlyTyrAspMetIleGlnAlaHisSerThrLeuGluHisAsnTyr 101
Qy 316 CTGCGCATGCTTTCTGTTCTCCGACGATGCTGAGGAGCTTCGATTCGCGGTCTTCC 375
Db 102 LeuGluAsnGlyPheTTrpAlaMetArgGluAlaAspGluLeuProAsnLysAspIleSer 121
Qy 376 GTGAAGTATCTGCCACATCGTGAAGTAAACGCAAGGTAGAGGACGACGCTCAAGGGT 435
Db 122 ValValPheIleAspLysIleGluAspValThrAspTyrValLysLysAspLeuLysAla 141
Qy 436 ATCACTGACGAGATGGAGCGCTCGGCAAGCTCAGGAGGTATGCCAAGAACTGCCCAA 495
Db 142 IleLysAspProAsnSerMetAspTyrLeuSerProLysTyrLeuGlnLysLeuAlaAsp 161
Qy 496 AAGAA-----AATGCACAGCAGAACTCTGCATC--GTAGAGCCT 537
Db 162 LysLysAlaGlyLysAsnPheSerAlaLysAsnProGlyLeuSerValGluIleLysAla 181
Qy 538 TTCATTCCAAACGAATACTTCTCATCTCATCGATGATTTCAGGAGCTTCGTATG 597
Db 182 PheTyrGlyGlyAsnLeuTyrLeuMetPheThrLysLysThrTyrThrAspValArgLeu 201
Qy 598 GTATTTGCTCTCCAGCTCTAGTAAAGTTCCGAGCGGATACGGACAACTGATGCG 657
Db 202 ValGlyAlaProThrSerIleGlyLysPheGlyAlaAspThrAspAsnTrpIleTrp 221
Qy 658 CCGGCTCACACGGCGCTTCAGGCTATTCGCTGATGCGGTGCGCGGACCAACGGCG 717
Db 222 ProArgHisThrGlyAspPheSerIlePheArgIleTyrAlaAspLysAsnGlyAsnPro 241
Qy 718 GCCGAATACAGCAGGACAAATAACCCCTATAAGCCGCTTACTTCGCTGCGGTATCCATG 777
Db 242 AlaProTyrSerGluAspAsnValProLeuLysProLysArgPhePheAsnIleSerLeu 261
Qy 778 CAAGCTACAGGCTGACGACTATGCCATGACCATCGGTTTCCGCGGACGACGATCGC 837
Db 262 GlyGlyValGlnGluAsnAspTyrAlaMetIleMetGlyPheProGlyThrHisArg 281
Qy 838 TACCTCACTCTCTGGGGTGTGGAAGATCGTATCGAAAACGAGAACAACTCCCTGATCGAA 897
Db 282 TyrPheThrAlaSerGluValAspLysTrpLysSerIleAspAsnAspIleArgIleArg 301
Qy 898 GTTCGCGGTATCAAGCAGGATCTGGAAGGACCGATGCGCGGACGATCAGGCTACCGGT 957
Db 302 MetArgAspIleArgGlnGlyValMetLeuArgGluMetLeuAlaAspProGlnIleLys 321
Qy 958 ATCAATATGCCAGCAACTGCTCAGAGTGTAACTATTGGAAGAACTTCGATCGGTATG 1017
Db 322 IleMetTyrSerAlaLysTyrAlaAlaSerGlnAsnAlaTyrLysArgAlaIleGlyAla 341
Qy 1018 AACCGCGTCTCGCTCTTTCAGCTGATAGTGTGCTGAAGCGTCCGAGGAAAGAGCATTC 1077
Db 342 AsnTrpAlaIleLysThrArgGlyLeuArgGlnAsnLysGlnAlaMetGlnAspArgLeu 361
Qy 1078 GCAGCTGGATCCCTAAGACGCGCAAGAGTGTCTCTATGCGGATGATGTTCTCTCTC 1137
Db 362 IleAlaTrpGlyAlaLysGlnGly---ThrProArgTyrGluGluAlaValHisGluIle 380
Qy 1138 GAAAGGCTTATAGGAGGAGGCAAGGCCCAACCGTCAGATGACTTATTTCAGGACGACG 1197
Db 381 AspAlaThrValAlaLysArgAlaAspLeuArgArgTyrTrpMetIleGluGlu 400
Qy 1198 CTCCTCGGTGCTACCGAGGTGGTTTCGT-----TTTGCACAGTTTTCACCAACGCA 1245
Db 401 IleIleArgGlyIleGluPheAlaArgSerProIleProThrGluAspGluThrLysAla 420
Qy 1246 TTGGCTACAATCTGATGCTCATGCGCGGTATCTCAATCGCTTGCAGCAAGTAC--- 1302
```

Db 421 LeuGlnGlyAsn-----AspAlaSerAla-----ArgLysGluAlaIleAspLysIleArg 437  
Qy 1303 -----AAAGACTACCTCCCTCGCTCGACCGTAAGTGTG 1338  
Db 438 ThrArgTyrSerLysPheAlaAsnLysAspTyrSerAlaGluValAspLysLysValAla 457  
Qy 1339 CCCGCCATGCTCGATATGTACGCCGGCGTATCCCTCCGACAGCTCCCGATATATTC 1398  
Db 458 ValAlaMetLeuThrGluTyrLeuLysGluIleProTyrGluAsnLeuPro---LeuHis 476  
Qy 1399 AAGAAATGTAATCGACAAAGATTCARAGGCGACGACGAGAGTATGCGATTCGTATTC 1458  
Db 477 LeuArgLeuValLysAspArgPheAlaGlyAspValGlnAlaTyrValAspAspIlePhe 496  
Qy 1459 GACAAGAGTGTGGTTCCTTATAGCAGCAAGATTCATGCCATGCTCAAGTCCATGCACAAG 1518  
Db 497 AlaArgSerValPheGlySerGluAlaGlnPheAspAlaPheAlaAlaValProSerVal 516  
Qy 1519 GAAAGTTTGCACAGGCTATCGGAGAAGATCCCGCAGTAGAGCTTCCACAGAGCTTAATA 1578  
Db 517 GluLysLeuAla-----GluAspProMetValLeuPheAlaSerSerValPhe 532  
Qy 1579 GCTGCTGCTCGCTATTCAGGCGGATGCGATGCCAATGCTATGCCATGAGAGGCG 1638  
Db 533 AspGluTyrArgLysLeuTyrAsnGluLeuArgProTyrAspAspProIleLeuAla 552  
Qy 1639 AAGCGTCTTTCTTTCGCGTTTTCGCTGAGATGTATCCCGGACGCTGCTCCGAGGCGAT 1698  
Db 553 GlnArgThrTyrIleAlaGlyLeuLeuGluMetAspGlyAspGlnAspGlnPheProAsp 572  
Qy 1699 GCCAATTCACCATGCTGATGAGTACGGTCCATCAAGGATATGAACCCGAGAGCGGT 1758  
Db 573 AlaAsnLeuThrLeuArgPheThrTyrGlyGlnValLysGlyTyrSerProArgAspAsn 592  
Qy 1759 GCGTGGTACAACTATCATACACAGGCAAGGCGCTATTGGAGAGCAGGATCCTTAAGAGC 1818  
Db 593 ValTyrTyrGlyHisGlnThrLeuAspGlyValMetGluLysGluAspProAspAsn 612  
Qy 1819 GATGAGTTTGCCTACAGGAGAAATATCTCGACCTCTTCCGACCAAAAACACTATGGTCG 1878  
Db 613 TrpGluPheValValAspProLysLeuLysAlaValTyrGluArgLysAspPheGlyArg 632  
Qy 1879 TATGCCGAG---ACGGTCAGCTCCATATCGCTTTCCTATCGACACAGCAGCATCAGGCG 1935  
Db 633 TyrAlaAspArgSerGlyArgMetProValAlaPheCysAlaThrHisThrThrGly 652  
Qy 1936 GGTAACTCCGCTAGCCCGTATTCGATAAGACGCGCTCTGTATCGCTGCTTCTCGAT 1995  
Db 653 GlyAsnSerGlySerProValMetAsnAlaAsnGlyGluLeuIleGlyLeuAsnPheAsp 672  
Qy 1996 GGCACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGACGCGCACATC 2055  
Db 673 ArgAsnTrpGluGlyValGlyAspIleGlnTyrLeuAlaAspTyrGlnArgSerIle 692  
Qy 2056 ACCGTGGACATCCGCTAGCTTCTTCTATGATTCGACAAATGGGTACGTGCGCCGCTC 2115  
Db 693 IleValAspIleArgTyrValLeuLeuValIleAspLysValGlyCysGlnArgLeu 712  
Qy 2116 ATCCAAGAGCTGAAGTTGATC 2136  
Db 713 LeuAspGluMetAsnIleVal 719

RESULT 4

US-10-008-355-7  
; Sequence 7, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355

; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 716  
; TYPE: PRT  
; ORGANISM: Xylella fastidiosa  
US-10-008-355-7  
  
Alignment Scores:  
Pred. No.: 8.02e-69 Length: 716  
Score: 907.50 Matches: 239  
Percent Similarity: 49.21% Conservative: 135  
Best Local Similarity: 31.45% Mismatches: 291  
Query Match: 23.76% Indels: 95  
Gaps: 21

US-10-008-355-1 (1-2139) x US-10-008-355-7 (1-716)

Qy 1 ATGCAAAATGMAATTAATAAAGATTATTCTTCGGAGCAGCCCTGCTGTTGGTGTCTCAGGG 60  
Db 1 MetArgPheAsnLeuLeuSerLeuSerValLeuAlaThrLeuIleThrValAspSer--- 19  
Qy 61 GTAGCCAAAGCCGACAAAGGCATGTGGCTCTCAACCAACTCAATCAGGAGAATCTGGAT 120  
Db 20 ---ThrHisAlaGlyGluGlyMetTrpValProGlnGlnLeu---ProGluIleAlaGly 37  
Qy 121 CGAATGCTGAGCTGCGGTTCACCTCCGTTGGATTGCTGCTACAGTTTCGACAAAGCCG 180  
Db 38 ProLysLysGlnAlaGlyLeuGlnLeuSerProGlnGlnLeuSerAsnLeuThrGlyAsp 57  
Qy 181 TCCATTGCCAATGCCGTGTTATCTTCGGTGGCGATACCGGTATCACAGTGTCCGAT 240  
Db 58 ProMetGlyAlaValValSerLeu-----GlyAsnCysThrAlaSerLeuValSerPro 75  
Qy 241 CAGGCGCTGATCTTTACCAACCACTGCGGATACGCTGCTATCCAGAGCCAAAGCAGC 300  
Db 76 GluGlyLeuValIleThrAsnHisCysAlaTyrGlyAlaIleGlnLeuAsnSerThr 95  
Qy 301 GTGATCAGCAGTATCTGCGCGATGTTTCTGTTCTCCGACGATGGTGAGAGCTTCGCG 360  
Db 96 ProLysLysAsnLeuIleLysGluGlyPheAsnAlaLeuThrGlnAlaAspGluValSer 115  
Qy 361 ATT---CGGGTCTTTCGCTCAAGTATCTGCGCAAGTACGTAAGGTAAACGACAAAGTA 417  
Db 116 AlaGlyProAsnAlaArgIleTyrValLeuGluGlnIleThrAspValThrAlaGlnAla 135  
Qy 418 GAAGGACAGCTCAAGGGTATCAGTACGAGATGGAGCGCTCTGCGCAAGCTCAGGAGGTA 477  
Db 136 LysAlaAlaLeuAlaAlaGlyAsnAspProPheLysArgThrThrAlaLeuGluThr 155  
Qy 478 TGC-----CAAGAACTGCGCCAAAAGAAAAATCGACGAGAGAACCACTCTGCTATC 531  
Db 156 PheSerLysGlnGluIleAlaLysCysGluGlu---GluGlnGlyTyrArgCysGlnPhe 174  
Qy 532 GAGCTTTCTTATTCACCAACGAAATCTCTCTACGCTACGATGTATTCAAGGAGCTT 591  
Db 175 PheSerPheAlaGlyGlyAsnThrTyrArgValPheLysAsnLeuGluIleLysAspVal 194  
Qy 592 CGTATGGTATTTGCTCTCCAGCTCTCTAGTAAAGTTTCGGAGCGATACGACGACACTGG 651  
Db 195 ArgLeuValTyrAlaProGlnGlySerValGlyLysPheGlyGlyAspValAspAsnTrp 214  
Qy 652 ATGTGGCGCGCTCACACGGGCGACTTCAGCTATTCCCGCTGTATGCCGCTGCCGACAAAC 711  
Db 215 MetTrpProArgHisThrGlyAspPheSerPheTyrArgAlaTyrValGlyLysAspGly 234  
Qy 712 CGGCGCGCGCAATACAGCAAGGACAAATAACCTATAAGCCGTTTACTTCTCGTCCCGTA 771  
Db 235 LysProAlaSerPheSerLysGluAsnIleProTyrArgProLysHisTrpLeuLysPhe 254

```
QY 772 TCCATGCAAGGCTACAAGGCTGACGACTATGACCATGACCATCGTTTCCGGGCGACTAGC 831
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 SerAspGlnProLeuGlyAspGlyAspPheValMetValAlaGlyTyrProGlyArgThr 274
QY 832 GATCGCTACTCCTCCTCTTGGGGTGGAGATCGTATCGAAACGAGAACCAATCCTCT 891
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 AsnArgTyrAlaLeuVal-----AlaGluPheGluAsnThrAlaHis 288
QY 892 ATCGAAGTTCGGGTATCAAGCAA-----GGCATCTGGAGGAGGCC 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 TrpThrTyrProValIleGlyGlnHisPheLysAsnLeuLeuAlaLeuIleGluAla 308
QY 934 ATGAGCGCAGATCAGGCTACCGGTATCAATATGCGCAGCAAGTATGCTCAGAGTGTAA 993
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 SerLysGlnAsnProAspIleGlnValLysTyrAlaSerThrLeuAlaGlyLeuAsnAsn 328
QY 994 TATGTGAAGAATTCGATCGGTATGAACCGCGTCTGCTGCTTGTACGTATAGTCTGT 1053
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 ThrSerLysAsnPheAspGlyGlnLeuAspGlyPheArgArgIleAsnAlaIleGlyGln 348
QY 1054 AAGCTGCGCGAGGAAGACATTCGACACTGGATCCGTAAGAACGCC----- 1101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 LysGlnSerGluGluThrAlaValLeuAlaTrpLeuLysGlnGlnGlyIleArgGlyHis 368
QY 1102 AAGAGTGTCTGTATGGCGATGATTGCTCTCTCGAAAGGCTTATAGGAAGGAGCC 1161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 GluAlaLeuAlaAlaHisGlnThrLeuValAspLeuThrGluGlnTyr----- 384
QY 1162 AAGGCCAAGCTGAGATGACTTATTTTGGCAGGACGCTCTCGGTGGTACCAGGTGGTT 1221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 LysAlaAsnGlnAspArgAspPheValLeuGlyGln--PheAsnGlySerGlyValIle 403
QY 1222 CGTTTGGCAGCTTTGCCAACCACTTGGCT-----ACAAATCCTGATCCTCAT--- 1269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 GlyValAlaValAsnLeuTyrArgIleAlaIleGluArgThrLysSerAspAlaGlnArg 423
QY 1270 -----GCCGGTATCCTCAAAATCGGTTGAC 1293
Db 424 GluAlaGlyTyrGlnGluArgAspLeuProThrIleGluGlyAsnLeuLysGlnMetGlu 443
QY 1294 GACAAGTACAAGACTACTCCCTCGCTCGACCGTAAGTG----- 1335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 -----ArgArgTyrLeuProGluMetAspArgGlnMetGlnGlnTyrTrpLeuThr 460
QY 1336 -----CTGCGCCCATGTCGATATTATGACCGCGCGCTATCCTGCC----- 1377
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 GluTyrAsnLysLeuPro-----ValLysGlnArgValAlaAlaIleAsp 475
QY 1378 -----GACAAGCTCCCGCATATATTCAAGAATGTAATCGACAAGAAATTCAAA 1425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 ValTrpLeuGlyAspGlyIleProAlaThrLeuLysArgLeuGlyAspThrLysLeuSer 495
QY 1426 GCGGACAGCAAGATATGACACTTGTATTCGACAGAGTGTGGTCTTATAGCGGAC 1485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 SerSerGluGluArg----- 500
QY 1486 AAGTTCATGCCATGCTCAAGTCCATGGACAGGAAAGTTTGCCAAAGGCTATCGAGAA 1545
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 -----LeuLysTrpPheAsnAlaAspArgAlaAlaPheGluSerSerGln 515
QY 1546 GATCGCGCAGTAGAGCTTTCCAAAGCGCTAATAGTCTGCTCGCGCTATTCAGGCGCGAT 1605
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 AspProAlaIleArgTyrAlaValAlaIleMetProAlaLeuLeuGluIleGluArgGln 535
QY 1606 GCGATGGCAATGCCATTCGCAATTCAGAGGCGACGCTCTTTCTTTCGCGGTTGGCT 1665
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 AsnLysIleArgThrGlyGluLeuLeuLysAlaArgProIleTyrLeuGlnAlaLeuAla 555
QY 1666 GAGATGTACCCC-----GGAGCTGCTCTCGCGAGCGATGCCAACTTCACCATGCGTATG 1719
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 AspTyrAsnLysSerHisGlyLysPheValTyrProAspAlaAsnSerSerLeuArgIle 575
QY 1720 AGCTACGGCTCCATCAAGGGATATGAACCGCAGGAGCGTGTGCTGTACAACTATCATACG 1779
```

```
Db 576 ThrPheGlyHisValLysGlyTyrSerProLysAspGlyValGluTyrThrProPheThr 595
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1780 ACAGGCAAGGCGTATTTGAGAGCAGGATCTCTAGAGCGGATGTTTCCCGTACAGGAG 1839
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 ThrLeuGlnGlyValMetAlaLys---AsnThrGlyValGluProPheAspSerProLys 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1840 ATATCTCTCGACCTCTTCCGACCAAAAATACTATGTCGCTATGCCGAGAAC----- 1893
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 SerLeuIleAsnAlaIleLysAlaLysSerTyrAlaAsnLeuAlaAspGlnArgIleGly 634
QY 1894 CAGCTCCATATCGCTTTCTCTATCGAACACGACATCACCGGCGGTAACTCCGGTAGCCCC 1953
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 ThrValProValAsnPheLeuSerAspLeuAspIleThrGlyLysSerGlySerPro 654
QY 1954 GTATTCGATAAGAGCGCGTCTGATCGGCTTGTGCTTCGATGGCACTGGGAAGCTATG 2013
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 ValLeuAspAlaHisGlyLysLeuValGlyLeuAlaPheAspGlyAsnTrpGluSerVal 674
QY 2014 ACTGGTGCATCGATTCGAACCCGATCTGCAGCGCACACATCAGCGTGCACATCCGCTAC 2073
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 SerSerAsnTrpValPheAspProValMetThrArgThrIleAlaValAspSerArgTyr 694
QY 2074 GTTCTCTTCATGATTCACAAATGGGTGTCAGTCCCGCTCTCATCCAGAGCTGAAGTTG 2133
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 695 ValGlnTrpIleMetThrGluValAlaProAlaProHisLeuLeuLysGluLeuAsnLeu 714
RESULT 5
US-10-008-355-5
; Sequence 5, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Bandula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235,00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; PRIORITY FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-5
Alignment Scores:
Pred. No.: 7,62e-67 Length: 734
Score: 884.50 Matches: 247
Percent Similarity: 48.58% Conservative: 129
Best Local Similarity: 31.91% Mismatches: 280
Query Match: 23.15% Indels: 118
Gaps: 25
US-10-008-355-1 (1-2139) x US-10-008-355-5 (1-734)
QY 6 AATGAATTAATAAGTATTCTTCGAGCAGCCCTGCTGTGGTGTCTTACGGGTAGC 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 AsnThrMetLysLysTrp-LeuLeuSerValAla-----ValAlaAlaSerPheAla 32
QY 66 CAAACCCGACAAAGCATGTGGCTCTCAACCAACTCAATCAGGAGATCTCGATCGAAT 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 32 rHisAlaAspGluGlyMetTrpGlnProHisGlnLeu---ProAlaMetAlaAspVal 51
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 GCGTCAGCTCGCGTTTACGCTCCCGTTGGATTTCGCTCTACAGTTCGCAAGCCCTCCAT 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 uLysAlaLysGlyLeuGluIleAspAlaLysSerIleSerLysLeuThrGluPhePro 71
QY 186 TGCCCAATCCGCTGGTGTATCTTCGCTGGCGGATGTACCGGTATCACAGTGTCCGATCAGG 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 71 t---AsnAlaValIleSerLeu---GlyGlyCysThrAlaSerPheValSerProLysG1 89  
QY 246 CCTGATCTTTACCAACCAACCACTCCGATACGGTGTCTATCCAGAGCAAGACGCGTGA 305  
Db 89 yLeuValValThrAsnHisHisCysAlaTyrglySerIleGlnTyrAsnSerThrProG1 109  
QY 306 TCAGCATATCGCGCATGTTTCGTTTCGTCGACCATGGGTGAGAGCTTCG---AT 362  
Db 109 uLysAsnLeuLeuGlnAspGlyPheLeuAlaLysThrPheAlaAspGluLeuProAlaL 129  
QY 363 TCCGGGTCTTCCCGTGAAGTATCTCGCGCAAGTCTGAAGGTAAAGGTGAAGG 422  
Db 129 aProGlySerArgValThrGluAspValThrAsnValThrGluArgValLysAl 149  
QY 423 ACAGCTCAAGGCTATCATCTACGAGATGGAGCGTCTCGCAAGCTCAGGAGGTATGCCA 482  
Db 149 aGlyLeuGluAsnLysThr-----GlyArgGluPheTyrgL 161  
QY 483 AGAAGTGGCCAAAAGAAATGCA-----GACGAGAACCACACT 521  
Db 161 nGlyValGluAsnGlnGluLysAlaLeuValAlaGluCysGluLysAspGluGlyTyrg 181  
QY 522 CTGCATCGTACAGCTTCTTATTCCAAACGAATACTTCTCATCTGCTACGATGTATT 581  
Db 181 gCysGlnValTyrSerPheHisGlyGlyLeuGluTyrrLeuValLysGlnLeuGluL 201  
QY 582 CAAGGAGCTTCGTATGTTTGTCTCCAGCTCTGTAGTAAAGTTTCGAGCGCATAC 641  
Db 201 eArgAspValArgLeuValTyrAsnProAlaGlySerValGlyLysTyrGlyGlyAspVa 221  
QY 642 GGCAACTGGATGTCGCGCTCACAGGGGACTTCAGCGTATTCGCGGTGTATGCGGG 701  
Db 221 lAspAsnTyrMetTyrProArgHisThrGlyAspTyrSerPheTyrgArgAlaTyrgL 241  
QY 702 TGCCGCAACCGCGCGCGCAATACAGCAAGCAATAAACCCTATAAGCCCGTTACTT 761  
Db 241 rLysAsnGlyLysProAlaGluPheSerAlaAspAsnValProTyrGluProLysSerPh 261  
QY 762 CGTGCCTGATCCATGCAAGCTCAAGGCTGAGCTATGCTATCCATGACCATCGTTTCCC 821  
Db 261 eLeuLysValSerAlaLysGlyLysValSerGluGlyAspPheValValAlaGlyTyrg 281  
QY 822 GGGCAGTACGATCGTACCTACCTCTTGGGGTGTGGAAGATCGTATCGAA-----AA 875  
Db 281 oGlyArgThrAsnArgTyrArgThrAlaThrGluValGlnAsnGluPheGluTyrAlaTy 301  
QY 876 CGAGAACAATCCTCGTGAAGTTCGCGGTATCAAGCAAGGCGATCTGGAAGGAAGCCAT 935  
Db 301 rProGluGlyLysMetLeuArgGluArgPheIleGlu-----IlelleLysAlaThrAl 319  
QY 936 GAGCGCAGATCAGGCTACCCGTATCAATATGCCAGCAAGTATGCTCAGAGTGTAACTA 995  
Db 319 aProGluGlySerAspGluArgIleLysTyrGluSerGlnIleAlaGlyLeuAlaAsnTy 339  
QY 996 TTGGAAGATTCGATCGGTATGAACCGCGGTCTCGCTGCTTTCGACGTGATAGTGTGTA 1055  
Db 339 rAlaLysAsnPheThrSerMetIleGluPheTyrgGlyLysSerThrMetLeuAlaAspAr 359  
QY 1056 CGGTGCCGAGGAGAGCATTCGAGACTGCGTAAAG---AACGGCAAGAGTGTGT 1112  
Db 359 gLysAlaLeuGluAlaLysLeuAlaGluTyrPileAlaLysAspSerArgGluAlaL 379  
QY 1113 CTATGGCATGATTTGTTCTTCTCGAAAGGCTTATAAGGAAGGAGCAAGGCAACCG 1172  
Db 379 sTyrgLysThrLeuAlaGluLeuAsp---AlaLeuIleAlaLysSerLysAlaHisG1 398  
QY 1173 TCAG-----ATGACTTATTTCAGCGACGCTCTTCGTTGCTGCTACCGAGGTGGT 1220  
Db 398 nGluArgAspMetIleLeuSerTyrrIleSerSerThrThrMetLeuProThr----- 415  
QY 1221 TCGTTTTTCACAGTTGTCACAGCATGGCTACAAATCCTCATGCTCATGCGGTATCCT 1280  
Db 416 -----AlaAsnAsnLeuTyrgArgLeu-----AlaHisGlu----- 425

QY 1281 CAAATCGCTTGACGACCAAGTACAAAGAC----- 1308  
Db 426 -LysGlnLeuProAspMetGlnArgGluProGlyPheGlnAspArgAspMetThrArgph 445  
QY 1309 -----TACCTCCCTCGCTCGACCTTAAGGTGCT 1337  
Db 445 eLysAlaSerMetGluArgIleAspArgArgTyrAlaAlaSerValAspLysAlaValLe 465  
QY 1338 GCCCGCATGCTCATATTTGACCGCGCGTATCCCTGCGCACAAAGCTCCCGCATATATT 1397  
Db 465 uPheAspMetLeuLysArgTyrAlaAlaLeuProGluAlaGlnArgLeuPro----- 482  
QY 1398 CAGAATGTATTCGACCAAGAAATTCAAAGCGCACACGAAG----- 1437  
Db 483 -----AlaMetAspLysAlaPheGlyIleAspAsnLysValAsnGluAlaLysLeuAl 500  
QY 1438 -AGTATGCGACACTTCGACAGAGTGTGGTTCCTTATAGCGAC---AAGTTCCA 1493  
Db 500 aLysThrLeuAspLysMetTyrAlaLysThrGluGlyAsnLysAspValArgLeuAl 520  
QY 1494 TGCCATGCTCAAGTCCATGGCAAGAAAGTTTGCCAAAGCTATCGAGAAAGATCGCGC 1553  
Db 520 aTrpMetLysSerValAspAspPheLys-----AlaSerLysAspProPh 536  
QY 1554 AGTAGAGCTTTTCCAGAGCGTA----- 1575  
Db 536 eIleGlnPheAlaValAlaMetTyrAspThrAsnMetSerGluGluLysLysGluLysG1 556  
QY 1576 -----ATAGCTGCTGCTCGCGCTATTCAGGCCCATGCGATGCGCATGCGCTA 1622  
Db 556 uLeuAspGlyGluLeuMetLysValArgProGlnTyrMetAspAlaIleIle---AlaTy 575  
QY 1623 TGCCATTGAGAAAGGCAAGCTCTTTCTTTCGCGTTTTCGCTGAGATGTACCCCGGACG 1682  
Db 575 rAsnLeuGluGlnGlyLysProValTyr----- 584  
QY 1683 TGCTCTGCCGCGCATGCCAACTTCACCATCGTATGAGTACGCTCCATCAAGGGATA 1742  
Db 585 -----AlaAspAlaAsnSerSerLeuArgValThrValGlyHisValLysGlyTy 601  
QY 1743 TGAACCGCAGGACCGTGCCTGGTACACTATCATACGACAGGCAAGGCGGTATTGGAGAA 1802  
Db 601 rSerProLysAspGlyLeuValAlaValProPheThrArgLeuGluGlyIleValGlnLy 621  
QY 1803 GCAGATCCTAAGACGATGAGTTTTCGCTACAGGAGAATATCTCGACCTCTTCCGAC 1862  
Db 621 s---AspThrGlyLysAspProPheAspAlaProLysGlnGlnLeuGluLysG1 640  
QY 1863 CAAAACATATGCTGCTATGCCGAGAAGCGT-----CAGCTCCATATCGCTTTCCTATC 1916  
Db 640 nLysGlnTyrGlyAspPheTyrrMetLysSerIleAspSerValProValAsnPheLeuSe 660  
QY 1917 GAACAACGACATCAGCGCGGTAACTCCGGTAGCCCGTATTCGATATGAGAACGCGCTCT 1976  
Db 660 rThrLeuAspThrThrGlyGlyAsnSerGlySerProThrLeuAsnGlyArgAlaGluLe 680  
QY 1977 GATCGGTCTTCTTCGATGCAACTGGGAAGCTATGAGTGGTGACATCGATTCGATTCGAACC 2036  
Db 680 uValGlyLeuLeuPheAspGlyValTyrGluSerIleIleGlyGlyTrpAlaPheAspAs 700  
QY 2037 CGATCTGCAGCGCAATACAGCGTGGACATCCGCTACGTTCTTCTCATGATTGACAAATG 2096  
Db 700 nGluIleAsnArgSerIleHisValAspSerArgTyrMetLeuTrpValMetLysTyrgLe 720  
QY 2097 GGGTCAGTCCCGCTCTCATCCAGAGCTCAAGTTGATC 2136  
Db 720 uAspHisAlaAspAsnLeuLeuAlaGluMetGluIleVal 733  
RESULT 6  
US-10-008-355-6  
; Sequence 6, Application US/10008355  
; Patent NO. US20020164759A1

GENERAL INFORMATION:  
APPLICANT: Travis, James  
APPLICANT: Potempa, Jan S  
APPLICANT: Banbula, Agnieszka  
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
FILE REFERENCE: 235.00440101  
CURRENT APPLICATION NUMBER: US/10/008.355  
CURRENT FILING DATE: 2001-11-08  
PRIOR APPLICATION NUMBER: US 60/246,827  
PRIOR FILING DATE: 2000-11-08  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 6

LENGTH: 732

TYPE: PRT

ORGANISM: Shewanella putrefaciens

US-10-008-355-6

Alignment Scores:

Pred. No.: 3,04e-60 Length: 732

Score: 807.50 Matches: 235

Percent Similarity: 44.81% Conservative: 119

Best Local Similarity: 29.75% Mismatches: 289

Query Match: 21.14% Indels: 147

DB: 9 Gaps: 25

US-10-008-355-1 (1-2139) x US-10-008-355-6 (1-732)

Qy 22 ATCTCTCGGAGAGCCCTCTGTTGGTGTCTCAGGGGTAGCCAAAGCCGACAAAGC 81

Db 3 IleAlaLeuValAlaAlaLeuValLeu---ThrCysGlyIleAlaThraAlaAspGluGly 21

Qy 82 ATGTGGCTCTCAAGCAACTCAATCAGGAGATCTGGATCGAATCGGTGAGCTGGCTTT 141

Db 22 GlnTrpGlnProTyrGlnMet---ProSerIleAlaAspLysLeuSerAlaArgGlyIle 40

Qy 142 AGCTCCGCTGGATGCTCTACAGTTTCGACAAAGCGTCCATGCGCAATGCCGTGTT 201

Db 41 AspileProAlaAspLysLeu-----AlaAspLeuThrSerTyrProMetAsnAlaVal 58

Qy 202 ATCTTCGGTGGGATGACCGGTATCATAGTCCGATCAGGCGCTCATCTTTACCAAC 261

Db 59 ValGlyLeuGlyTyrCysThrAlaSerPheValSerProGlnGlyLeuValValThrAsn 78

Qy 262 CACCACCTCGGATCGGTCTATCCAGAGCCAAAGCGGTGGATCAGCACTATCTCGGC 321

Db 79 HisHisCysAlaTyrLysAlaIleGlnTyrAsnThrLysLysGluHisAsnTyrLeuGlu 98

Qy 322 GATGGTTTCGTTTCGCGCAGGATGGGTGAGGAGCTTCGATTCGGGTCTTTCCGTGAAG 381

Db 99 GlnGlyPheLeuAlaThrSerMetAspLysGlu---ProSerAlaGlyProAsnGluArg 117

Qy 382 ---TATCTCGCCAGATCGTGAAGTACGACAGCAAGGTAGAGGACGCTCAAGGGTATC 438

Db 118 LeuTyrIleThrGlu-----AlaValThrAspValThrSerAspValThrLysAspLeu 135

Qy 439 ACT---GACGAGATGGAGCTGCGCAAGAGCTCAGGAGGTATGCCAAAGCTGCCAAA 495

Db 136 SerGlnAspProLeuLysArgTyrGluGluIleGluAsnHisSerLysAlaLeuIleLys 155

Qy 496 AAGAAATGCGAGGAGAACCACTCTGCATCTGATGAGCCCTTCTATTCACCAACGAA 555

Db 156 SerCysGluAlaAspAspAsnTyrArgCysAsnValArgSerPheHisAsnGlyLeuGlu 175

Qy 556 TACTCTCTCATCGTACGATGATTCAAGGACGCTGCTATGTTGCTTCTCTCCCGCAGC 615

Db 176 TyrTyrLeuIleLysGlnLeuMetIleArgAspValArgLeuValTyrAlaProProGlu 195

Qy 616 TCTGTAGTAAGTTCGGAGGCGATACGACAACTGGATGTGGCGCGTCACACGGCGCAC 675

Db 196 SerValGlyGlyTyrGlyGlyAspileAspAsnTyrGluTyrProArgHisSerGlyAsp 215

Qy 676 TTCAGCGTATTCCCGGTGTATGCGCGGTGCCGACAAACCGCGCGCGGCGAATACAGCAAGGAC 735

Db 216 PheAlaPheLeuArgAlaTyrValGlyLysAspGlyLysProAlaAlaPheSerGluAsp 235

Qy 736 AATAAACCTATATAGCCCGCTTACTTCCTCGCTATCCATGCAAGGCTACAAAGCTGAC 795

Db 236 AsnIleProTyrThrProLysSerTyrLeuLysValAsnAlaAspGlyValLysAlaGly 255

Qy 796 GACTATGCCATGACCATCGGTTTCCCGGCGAGTACGGATCGCTAC---CTCAGTTCT--- 849

Db 256 AspGlyValPheValAlaGlyTyrProGlyThrThrAsnArgTyrAsnLeuThrSerGlu 275

Qy 850 -----TGGGGTGTGGAAGATCGTATCTGAAACAGAGAACAACTCTCTCGT 891

Db 276 LeuLysPheAlaSerAspTrpLeuTyrProThrGlnAlaLysArgTyr-----Gln 292

Qy 892 ATCGAAGTTCGGGTATCAACAGGATCTGGAAGGAGCAATG---AGCGCAGATCAG 948

Db 293 LeuGlnIleAspThrIle-----GluAlaMetGlyGlnLysAspAla 306

Qy 949 GCTACCCGCTATCAATATGCCAGCAAGTATGCTCAGAGTGTCTAATATTGGAAGAAATTCG 1008

Db 307 AspileAlaIleLysTyrAlaGlyAsnMetAlaSerMetAlaAsnArgMetLysLysLeu 326

Qy 1009 ATCGGTATGAACCGGCTCTCGCTCTTTCAGCTGATAGTTCGTAAGCGTCCGAGGAA 1068

Db 327 AsnGlyLeuLeuAlaGlyPheLysAlaThrAspileValGlyIleLysGlnGlnArgGlu 346

Qy 1069 AGACATTCGCGAGACTGGATCCGTAAGACGGCAAG----- 1104

Db 347 AsnAspPheLeuAlaTrpLeuThrLysAsnProAsnLeuAsnGlnAsnLeuIleSerGlu 366

Qy 1105 -----AGT 1107

Db 367 LeuGluValLeuLeuAlaGluGlnGlnLeuGlnThrGlnThrAsnTyrTyrPheThrAsn 386

Qy 1108 GCTGTCTATGCGGATGATTCCTCTCTCGAAAAGGCTTAT-----AAG 1152

Db 387 AlaGlnSerSerThrLeuLeuThrAlaAlaAsnAsnLeuTyrArgLeuAlaLysGluLys 406

Qy 1153 GAAGGAGCAAGCAACCGTGCAGTACTTATTGACGAGACGCTCTTCGGTGGTACC 1212

Db 407 GlnLysSerAspAlaGluArgGluIleGlyTyrGlnGluArgAspLeu----- 422

Qy 1213 GAGGTGGTTCGTTTGCACAGTTTCCCAACGATTTGGCTACAAATCTGTATGCTCATGCC 1272

Db 423 -----AlaMetPheSerSerArg----- 428

Qy 1273 GGTATCTCAATCGCTTGACGACAAAGTACAAAGACTACCTCCCTCGCTCGACCGTAAG 1332

Db 429 -----LeuLysArgIleAspSer-----SerPheAspValLys 439

Qy 1333 GTG-----CTGCCCGCCATGCTCGATATGTCAGCGCGCGT 1368

Db 440 ValAspLysThrLeuTrpLeuGlnAspLeuAsnAlaTyrLeuSerGlnProAsnArgVal 459

Qy 1369 ATCCCTGCGCGACAGCTCCCGATATATTCAAGAAATGTAATCGACAAAGAAATTCAAAGGC 1428

Db 460 AlaAlaLeuAspAsnMet-----LeuAsnLeuAsnAspLys----- 471

Qy 1429 GACAGGAAGAGTATGACAGCTTCGTATTCGACAGAGAGTGTGGTTCCTTATAGGACAAAG 1488

Db 472 -----AsnValSerLeuAlaAlaLysLeuAspGly 481

Qy 1489 TTCCATGCTATGCTCAAGTCCATCCAGCAAG-----GAAAAGTTT 1527

Db 482 LeuTyrSerLeuThrLeuThrAspGlnAlaGlnArgLeuAlaTrpMetGluAlaAsp 501

Qy 1528 GCCAAGGCTATCGAG-----AAAGATCCGCGAGTAGAGCTTTCCAAAGAGCGTAATAGCT 1581

Db 502 AlaLysAlaPheGluThrSerSerAspProPheIleArgLeuAlaValAlaLeuTyrAsp 521

Qy 1582 GCTGCTCGCGGTATTCAGGCGCGATGCGCAATGCGCTATGCTATGCAAGGCGCAAG 1641

```
Db 522 ThrAsnMetAlaGlnGluLysAlaGluLysIleLeuAlaGlyLysLeuSerThrAlaArg 541
QY 1642 CGCTTTTCCCGTTTCGGTGTGAGATGATACCCCGAGCGTCTCTGCG-----AGC 1695
Db 542 ProAlaTyrMetAlaAlaValIleAspTyrTyrLysAlaAsnAsnTrpProValTyrPro 561
QY 1696 GATGCCAATTCACCATCGGTATGAGTACCGGTCTCCATCAAGGATATGAACCGCAGGAC 1755
Db 562 AspAlaAsnGlyThrLeuArgIleSerTyrGlyMetValAspGlyTyrGlnSerArgAsp 581
QY 1756 GGTGCCTGTACAACTATCATACGACGAGCAAGCGCTATTGGAGAGCAGGATCCTAAG 1815
Db 582 AlaLeuTyrLysGlnProPheThrArgLeuAspGlyIleValAlaLysHis---ThrGly 600
QY 1816 AGCGATGAGTTTCCCGTACAGGAGATATCTCCGACCTCTTCCGCACCAAAACTATGTT 1875
Db 601 ValGluProTyrAsnAlaProLysLysLeuLeuAspAlaIleSerValGlnArgPheGly 620
QY 1876 CGCTAT----- 1881
Db 621 AspHisLeuValLysSerValTyrGlnAspProArgGlyTrpIleCysArgLeuPheSer 640
QY 1882 -----GCCGAGAAGGTCAGCTCCATATCCCTTCTCTATCGAACACGAC 1926
Db 641 CysLeuAspLysProGluGluPheAsnSerValProValAsnPheLeuSerSerValAsp 660
QY 1927 ATCAGCGCGGTAACTCCCGTAGCCCGTATTCGATAAGAAGCAGCGCTCTGATCGGTCTT 1986
Db 661 ThrThrGlyGlyAsnSerGlySerProValPheAsnGlyLysGlyGluLeuValGlyLeu 680
QY 1987 GCTTTCGATCGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAG 2046
Db 681 AsnPheAspSerThrTyrGluAlaIleThrLysAspTrpPhePheAsnProThrIleThr 700
QY 2047 CGCACATACGCTGACATCCGCTAGCTTCTCTCATGATTGACAAATGGGTGCTAGTGC 2106
Db 701 ArgAlaValHisValAspIleArgTyrIleLeuTrpMetMetAspGluValAspHisAla 720
QY 2107 CCCTGCTCATCAAGAGCTGAAGTTGATC 2136
Db 721 AspAsnLeuIleLysGluLeuAspLeuVal 730
RESULT 7
US-10-008-355-3
; Sequence 3, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-3
Alignment Scores:
Pred. No.: 3 62e-15 Length: 52
Score: 274.00 Matches: 52
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7.17% Indels: 0
Gaps: 0
US-10-008-355-1 (1-2139) x US-10-008-355-3 (1-52)
```

```
QY 1930 ACGGGCGGTAACTCCGCTAGCCCGCTATTGATAAAGAACGCCCTCTGATCGGTCTTGCT 1989
Db 1 ThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla 20
QY 1990 TTCGATGGCAACTGGGAGCTATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGC 2049
Db 21 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 40
QY 2050 ACAATCAGCGTGGACATCCGCTAGCTTCTCTTCATG 2085
Db 41 ThrIleSerValAspIleArgTyrValLeuPheMet 52
RESULT 8
US-09-945-917-3
; Sequence 3, Application US/09945917
; Publication No. US20030042381A1
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/945,917
; CURRENT FILING DATE: 1998-09-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1528
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-945-917-3
Alignment Scores:
Pred. No.: 0.00144 Length: 1528
Score: 148.50 Matches: 106
Percent Similarity: 37.48% Conservative: 84
Best Local Similarity: 20.91% Mismatches: 211
Query Match: 3.89% Indels: 107
Gaps: 19
US-10-008-355-1 (1-2139) x US-09-945-917-3 (1-1528)
QY 226 ATCAGCTGTCCGATCAGCGCTGATC-----TTTACCAACACCACCTGC 270
Db 1 MetThrThrSerAsnValGluLeuIleProIleTyrThrAspTrpAlaAsnArgHisLeu 20
QY 271 GGATACGGTGTCTATC---CAGAGCCAAAGCAGCGGTGGATCAGGATATCTCCGCGATGGT 327
Db 21 SerLysGlySerLeuSerLysSerIleArgAspIleSerAsnAspPheArgAspTyrArg 40
QY 328 TTCGTTTCTCGCAGCATGGGTGAGGACCTTCCGATTCGGGTCTTTCGCTGAAGTATCTG 387
Db 41 LeuValSerGlnLeuIleAsnValIleValProIleAsnGluPheSerProAlaPheThr 60
QY 388 CGCAAGATCGTGAAGTAACGCAAGTAGAAGCAGCTCAAGGCTATCAGGTATCAGTACGAG 447
Db 61 LysArgLeuAlaLysIleThrSerAsnLeuAspGly---LeuGluThrCysLeuAspTyr 79
QY 448 ATGAGCGCTCTGCGCAAGAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAGAAATGCA 507
Db 80 LeuLysAsnLeuGlyLeuAsp-----CysSerLysLeuThrLysThrAspIleAsp 96
QY 508 GACGAGAACCACTCTCGATCGTAGACCCCTTCTATTCCAAACAGCAACTACTTCTCTCATC 567
Db 97 SerGlyAsnLeuGlyAlaValLeuGlnLeuLeuPheLeuLeuSerThrTyrLysGlnLys 116
QY 568 GTCCTACCATGATTCAAGGACGTTCTGATGTATTTGCTCTCCCTCCAGCTCTGTAGGTAAG 627
Db 117 LeuArgGlnLeuLysLysAspGlnLysLysLeuGlnLeuProThrSerIle-MetPr 136
```

```
QY 628 TTCGGAGCGGATACGGACAACATGGATGTGGCCGCGTCACACGGCGACTTCAGCGTATTC 687
Db 136 oProAlaValSerLys----- 141
QY 688 CGCGTGATGCGCGTCCGACACACCGCGCGCGAATACACGAGGACAAATAAACCCCTAT 747
Db 142 -----LeuProSerProArgValAlaThrSerAlaThrAlaSerAlaThrAsnProAs 159
QY 748 AAGCCCGTTTACTTCGTGCGGTATCCATCAAGGCTCAAGGCTGACGAGCTATGCGCATG 807
Db 159 nSerAsnPheProGlnMet-----SerThrSerArgLeuGlnThr----- 172
QY 808 ACCATCGTTTCCGGCGCAGTACGATCGCTACTCTCTTGGGGTGTGGAAGATCGT 867
Db 173 ----ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyIleLysProLysTh 191
QY 868 A-----TCGAAACGAGACAATCCTCGTATCGAAGTTCCG 903
Db 191 rSerGlyLeuLysProProSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211
QY 904 GGTATCAAGCAAGCATCTGA----- 925
Db 211 gProSerSerArgSerSerGlyAsnAsnValGlySerThrIleSerThrSerAlaLy 231
QY 926 -----AGGAAGCCATGAGCGCA 942
Db 231 sSerLeuGluSerSerSerThrTyrSerSerIleSerAsnLeuAsnArgProThrSerGl 251
QY 943 GATCAGGCTACCGTATCAATATGCGCAGCAAGTATGCTCAGAGTGCTACTATTGAAG 1002
Db 251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrLy 270
QY 1003 AATTCGATGCT-----GTATGAACCGCGTCTCGCTCTGCTTACGTCATAGGT 1050
Db 270 sIleGlySerSerLysLeuAlaAlaProLysAlaValSerThrProLysLeuAlaSerVa 290
QY 1051 CGTAAGCGTCCGAGGAAGAGCATTCGCGAGACTGGATCCGTAAGAAGCGCAAGAGTCT 1110
Db 290 lLysThrIleGlyAlaLysGlnGluProAspAsnSerGlyGlyGlyGlyMetLe 310
QY 1111 GTCTATGCGGATGTATGCTTCTCTCGAAAGGCTTATAGGAGGAGCGCAAGGCCAAC 1170
Db 310 uLysLeuLysLeuPheSerSerLysAsnProSerSerSerSerSerProGlnProTh 330
QY 1171 CGTCAGATGACTTATTTGAGCGAGCACTCTTCGCTGCTACCGAGGTTGCTGTTTGA 1230
Db 330 r-----ArgLysAlaAlaValProGlnGlnThrLeuSe 343
QY 1231 CAGTTTCCCAACGATTTGGGTACAAATCTCTGATGCTCATCGCGGTATCTCAATCGCTT 1290
Db 343 rLysIleAlaAla-----ProValLysSerGlyLeuLy 354
QY 1291 GACGACAAGTACAAGACTACCTCCCTCGCTCGACCGTAAAGTGTGCGCGCATGCTC 1350
Db 354 sProProThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373
QY 1351 GATATTGTAGCGCGGTATCCCTCGCGACCAAGCTCCCGCATATATCAAGATGTAATC 1410
Db 373 sValSerTyrArg-----LysThrAspAlaProIleSerGlnGlnAspSe 389
QY 1411 GACAAGAAATTCAAAGCGACACGAAGAAG-----TATGACAGACTTCGTATTCGACAA 1463
Db 389 rLysArgCysSerLysSerSerGluGluSerGlyTyrAlaGlyPhe-----As 406
QY 1464 GAGTGTGTTCTTATAGGCACAGTTCCATGCCATGCTCAAG---TCCATGGACAAGGA 1520
Db 406 nSerThrSerProThrSerSerSerThrGluGlySerLeuSerMetHisSerThrSerSe 426
QY 1521 AAAGTTTCCCAAGGCTATCGAAGAAGATCCGGCA-----GTACAGACTTTCACGAAG 1571
Db 426 rLysSerSerThrSerAspGluLysSerProSerSerSerSerSerSerThrLeuAsnAlaSe 446
QY 1572 CGTAATAGTCTGCTCGC 1590
```

```
Db 446 rIleValThrAlaIleArg 452
RESULT 9
US-09-945-917-4
; Sequence 4, Application US/09945917
; Publication No. US20030042381A1
; GENERAL INFORMATION:
; APPLICANT: Bogaert, Thierry
; APPLICANT: Vandekerckhove, Joel
; TITLE OF INVENTION: PROCESSES FOR THE IDENTIFICATION OF COMPOUNDS WHICH
; TITLE OF INVENTION: CONTROL CELL BEHAVIOUR, THE COMPOUNDS IDENTIFIED AND
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING THEM AND THEIR
; TITLE OF INVENTION: USE IN THE CONTROL OF CELL BEHAVIOUR
; FILE REFERENCE: P/14-1
; CURRENT APPLICATION NUMBER: US/09/945,917
; CURRENT FILING DATE: 1998-09-21
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1583
; TYPE: PRF
; ORGANISM: Caenorhabditis elegans
US-09-945-917-4
Alignment Scores:
Pred. No.: 0.00147 Length: 1583
Score: 148.50 Matches: 106
Percent Similarity: 37.48% Conservative: 84
Best Local Similarity: 20.91% Mismatches: 211
Query Match: 3.89% Indels: 107
DB: Gaps: 19
US-10-008-355-1 (1-2139) x US-09-945-917-4 (1-1583)
QY 226 ATCACAGTGTCCGATCAGGCGCTGATC-----TTTACCAACCAACCACTGC 270
Db 1 MetThrThrSerAsnValGluLeuIleProIleThrThrAspTrpAlaAsnArgHisLeu 20
QY 271 GGATACAGTCTATC---CAGAGCCAAAGCAGCGTGCATCAGACTATCTCGCGATGGT 327
Db 21 SerLysGlySerLeuSerLysSerIleArgAspIleSerAsnAspPheArgAspTyrArg 40
QY 328 TTCCTTTCTCGCAGATGGGTGAGGAGTTCCTCGATTCCTCGGTCTTTCCGTAAGTATCTG 387
Db 41 LeuValSerGlnLeuIleAsnValIleValProIleAsnGluPheSerProAlaPheThr 60
QY 388 CGCAAGATCGTGAAGTAGACGACAGGTAAGAGCAGCTCAAGGATATCATCTAGCAGAG 447
Db 61 LysArgLeuAlaLysIleThrSerAsnLeuAspGly---LeuGluThrCysLeuAspTyr 79
QY 448 ATGAGAGCTGTGCGCAAGCTCAGGAGTATGCCAAGACTGGCCAAAGAAATAATGCA 507
Db 80 LeuLysAsnLeuGlyLeuAsp-----CysSerLysLeuThrLysThrAspIleAsp 96
QY 508 GACGAGAACCAACTCTGCATCGTAGAGCTTTCTATTCCAAACAAAGTAATCTTCTCATC 567
Db 97 SerGlyAsnLeuGlyAlaValLeuGlnLeuLeuLeuLeuSerThrTyrLysGlnLys 116
QY 568 GTCGTACGATGTATTCAGAGCGTTCGTATGTTATTTCTCTCTCCAGCTCTGTAGGTAA 627
Db 117 LeuArgGlnLeuLysLysAspGlnLysLeuGlnLeuProThrSerIle-MetPr 136
QY 628 TTCGAGGCGATACGGACAACACTGGATGTCGCGCGTCACACGGCGACTTCAGCGTATTC 687
Db 136 oProAlaValSerLys----- 141
QY 688 CGCGTGATTCGCGGTGCGGACACCGCGCGCGCGCAATACAGACAGCAATAAACCCCTAT 747
Db 142 -----LeuProSerProArgValAlaThrSerAlaThrAlaSerAlaThrAsnProAs 159
QY 748 AAGCCCGTTTACTTCGTGCGGTATCCATCAAGGCTCAAGGCTGACGAGCTTCAGCGTATTC 807
Db 159 nSerAsnPheProGlnMet-----SerThrSerArgLeuGlnThr----- 172
QY 808 ACCATCGTTTCCGGCGCAGTACGATCGCTACTCTCTTGGGGTGTGGAAGATCGT 867
Db 173 ----ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyIleLysProLysTh 191
QY 868 A-----TCGAAACGAGACAATCCTCGTATCGAAGTTCCG 903
Db 191 rSerGlyLeuLysProProSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211
QY 904 GGTATCAAGCAAGCATCTGA----- 925
Db 211 gProSerSerArgSerSerGlyAsnAsnValGlySerThrIleSerThrSerAlaLy 231
QY 926 -----AGGAAGCCATGAGCGCA 942
Db 231 sSerLeuGluSerSerSerThrTyrSerSerIleSerAsnLeuAsnArgProThrSerGl 251
QY 943 GATCAGGCTACCGTATCAATATGCGCAGCAAGTATGCTCAGAGTGCTACTATTGAAG 1002
Db 251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrLy 270
QY 1003 AATTCGATGCT-----GTATGAACCGCGTCTCGCTCTGCTTACGTCATAGGT 1050
Db 270 sIleGlySerSerLysLeuAlaAlaProLysAlaValSerThrProLysLeuAlaSerVa 290
QY 1051 CGTAAGCGTCCGAGGAAGAGCATTCGCGAGACTGGATCCGTAAGAAGCGCAAGAGTCT 1110
Db 290 lLysThrIleGlyAlaLysGlnGluProAspAsnSerGlyGlyGlyGlyMetLe 310
QY 1111 GTCTATGCGGATGTATGCTTCTCTCGAAAGGCTTATAGGAGGAGCGCAAGGCCAAC 1170
Db 310 uLysLeuLysLeuPheSerSerLysAsnProSerSerSerSerSerProGlnProTh 330
QY 1171 CGTCAGATGACTTATTTGAGCGAGCACTCTTCGCTGCTACCGAGGTTGCTGTTTGA 1230
Db 330 r-----ArgLysAlaAlaValProGlnGlnThrLeuSe 343
QY 1231 CAGTTTCCCAACGATTTGGGTACAAATCTCTGATGCTCATCGCGGTATCTCAATCGCTT 1290
Db 343 rLysIleAlaAla-----ProValLysSerGlyLeuLy 354
QY 1291 GACGACAAGTACAAGACTACCTCCCTCGCTCGACCGTAAAGTGTGCGCGCATGCTC 1350
Db 354 sProProThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373
QY 1351 GATATTGTAGCGCGGTATCCCTCGCGACCAAGCTCCCGCATATATCAAGATGTAATC 1410
Db 373 sValSerTyrArg-----LysThrAspAlaProIleSerGlnGlnAspSe 389
QY 1411 GACAAGAAATTCAAAGCGACACGAAGAAG-----TATGACAGACTTCGTATTCGACAA 1463
Db 389 rLysArgCysSerLysSerSerGluGluSerGlyTyrAlaGlyPhe-----As 406
QY 1464 GAGTGTGTTCTTATAGGCACAGTTCCATGCCATGCTCAAG---TCCATGGACAAGGA 1520
Db 406 nSerThrSerProThrSerSerSerThrGluGlySerLeuSerMetHisSerThrSerSe 426
QY 1521 AAAGTTTCCCAAGGCTATCGAAGAAGATCCGGCA-----GTACAGACTTTCACGAAG 1571
Db 426 rLysSerSerThrSerAspGluLysSerProSerSerSerSerSerSerThrLeuAsnAlaSe 446
QY 1572 CGTAATAGTCTGCTCGC 1590
```



```
Db 159 nSerAsnPheProGlnMet-----SerThrSerArgLeuGlnThr----- 172
QY 808 ACCATCGTTTCCGGCAGCAGTACGATCGTACTCTCTTGGGTGGTGAAGATCGT 867
Db 173 ----ProGlnSerArgIleSerLysIleAspSerSerLysIleGlyIleLysProLysTh 191
QY 868 A-----TCGAAACAGCAGAACAACTCTCGTATCGAAGTTTCG 903
Db 191 rSerGlyLeuLysProProSerSerSerThrThrSerSerAsnAsnThrAsnSerPheAr 211
QY 904 GGTATCAAGCAGCATCTCGA----- 925
Db 211 gProSerSerArgSerSerGlyAsnAsnValGlySerThrIleSerThrSerAlaLy 231
QY 926 -----AGGAAGCCATCAGCGCA 942
Db 231 sSerLeuGluSerSerSerThrThrTySerSerIleSerAsnLeuAsnArgProThrSerGl 251
QY 943 GATCAGGCTACCGGTATCAATATGCGCAGCAAGTATGCTCAGAGTGTAACTATTGGAAG 1002
Db 251 nLeuGlnLysPro---SerArgProGlnThrGlnLeuValArgValAlaThrThrThLy 270
QY 1003 AATTCGATCG-----GTATCAACCGCGGTCTCGTCTGTGACGTATAGGT 1050
Db 270 sIleGlySerSerLysLeuAlaLaProLysAlaValSerThrProLysLeuAlaSerVa 290
QY 1051 CGTAAGCGTCCGAGGAAAGAGCATTCGCCAGACTGGATCGCTAAGACGGCAAGATGCT 1110
Db 290 lLysThrIleGlyAlaLysGlnGluProAspAsnSerGlyGlyGlyGlyMetLe 310
QY 1111 GTCTATGGGATGATTGTCTCTCGAAAGGCTTATAGGAAGGAGCGCAAGCCAA 1170
Db 310 uLysLeuLysLeuPheSerSerLysAsnProSerSerSerAsnSerProGlnProTh 330
QY 1171 CGTGAGATGACTATTATGACGGAGCACTCTCGGTGTACCGAGGTGTCGTTTTCGA 1230
Db 330 f-----ArgLysAlaAlaAlaValProGlnGlnThrLeuSe 343
QY 1231 CAGTTTCCCAACGATGGTACAAATCCTGATGCTCATCGCGGTATCCTCAAAATCGCTT 1290
Db 343 rLysIleAlaAla-----ProValLysSerGlyLeuLy 354
QY 1291 GACGACAAGTACAAGACTACCTCCCTCGGTGACCGTAAAGGTGTCGCGCGCATGCTC 1350
Db 354 sProProThrSerLysLeuGlySerAlaThrSerMetSerLysLeuCysThrPro---Ly 373
QY 1351 GATATTCTAGCGCGCTATCCCTCGCGCAGCAAGTCCCGCATATATTCAAGAATGTAATC 1410
Db 373 sValSerTyArg-----LysThrAspAlaProIleLeSerGlnGlnAspSe 389
QY 1411 GACAAGAAATTCAAAGCGACACCAAGAAAG-----TATGCACTCTCGTATTTCGACAA 1463
Db 389 rLysArgCysSerLysSerSerGluGluGluSerGlyTyAlaGlyPhe-----As 406
QY 1464 GAGTGTGGTCTTATAGCAGCAAGTTCCATGCCATGCTCAAG---TCCATGGACAGGA 1520
Db 406 nSerThrSerProThrSerSerThrThrGluGlySerLeuSerMetHisSerThrSerSe 426
QY 1521 AAAGTTTCCCAAGCTATCAGCAAGATCCGCGCA-----CTAGAGCTTTCAGAG 1571
Db 426 rLysSerSerThrSerAspGluLysSerProSerSerAspAspLeuThrLeuAsnAlaSe 446
QY 1572 CGTAATAGCTGCTGCTCGC 1590
Db 446 rIleValThrAlaIleArg 452
```

## RESULT 10

```
US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
```

```
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: NO. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801.368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: PR1
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

Alignment Scores:
Pred. No.: 0.0027 Length: 1367
Score: 145.00 Matches: 160
Percent Similarity: 33.69% Conservative: 95
Best Local Similarity: 21.14% Mismatches: 354
Query Match: 3.80% Indels: 148
DB: 25

US-10-008-355-1 (1-2139) x US-09-801-368-108 (1-1367)
QY 65 CCAAGCCGACAAAGCATGTGGCTCCTCAACGAACTCAATCAGAGAAATCTGGATCGAA 124
Db 355 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 374
QY 125 TGGTGAGCTCGGCTTTACGCTCCGTTGGATTGCTCTACAGTTTCGACAAAGCGTCCA 184
Db 375 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerAlaPro 394
QY 185 TTGCAATCGCTGTATCTTCCGTG-----CGGATGTACCGGTATCACAGTGTCCG 238
Db 395 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThr 414
QY 239 ATCAGGCGCTGATCTTTACCAACCCACCCTCGGATACGGTGTCTATCCAGACCAAGCA 298
Db 415 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 434
QY 299 CGGTGGATCAGCACTATCTCGCGCATGTTTCTTCCGACCATGGTGGTGGAGGCTTC 358
Db 435 ThrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSer 454
QY 359 CGATTCCGGGTC---TTTCCGTGAAGTATCTGCGCAAGATCGTGAAGTAAAGGACAAG 415
Db 455 SerAlaProValProThrProSerSerSerThrThrGluSer----- 468
QY 416 TAGAAGGACAGCTCAAGGGTATCAGTACGAGATGGAGCTGCGCAAGACTCAGGAGG 475
Db 469 ---SerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProValProThr 487
QY 476 TATCCCAAGAACTGGCCCAAAAAAAGAAATGACAGGAGAACCACTCTGCTAGTAGAGC 535
Db 488 ProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSer 507
QY 536 CTTTCTATCCCAACAGCAATACCTCCTCATCGTCTACGATGTATTCAAGGACGTTCCGTA 595
Db 508 SerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAla----- 525
```

Qy 596 TGGTATTTGCTCCCTCCAGCTCTGTAGTAAGTTGGAGGCGATACGGACAACTGGATGT 655  
Db 526 -----ProAlaProThrProSerSerThrThrGluSerSerSerAla 540  
Qy 656 GGCCTGTCACACGGCGACCTACCGTATTTCCGGCTGTATCCGGTCCGCCAACACCGGC 715  
Db 541 ProValThrSerSerThrThrGluSerSerAla-----ProValProThrProSer 558  
Qy 716 CGGCGAATACAGCAAGACATAACCCCTATAAGCCGCTTACT-----TCG 763  
Db 559 SerSerThrThrGluSerSerSerThrProValThrSerSerThrThrGluSerSerSer 578  
Qy 764 CTGCCGTATCCA-----TGCAAGGCTACAAGGCTAGCAGCTATCCATGACCATCG 814  
Db 579 AlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThr 598  
Qy 815 GTTTCGGGGCAGTACGGATCGCTACCTACCTCTTCTGGGGTGTGGAAGATCGTATCGA 874  
Db 599 ProSerSerSerThrThrGluSerSerSerAlaProAlaProThrProSerSerSerThr 618  
Qy 875 ACGAGAACATCTCGTATCAAGTTCGCGGTATCAAGCAAGGATCTGGAAGGAGCCA 934  
Db 619 ThrGluSerSerSerAlaProValThrSerSerSerThrThrGluSerSer----- 635  
Qy 935 TGAGCGCAGATCAGGCTACCGGTATCAAAATATGCCAGCAAGTATGCTCAGAGTGTAACT 994  
Db 636 ---AlaProValProThrProSerSerSer-----ThrThr 646  
Qy 995 ATTGGAAGAAATTCGATCGGTATGAACCGGCTCGCTCGCTCTGACGTATAGTGTCTGA 1054  
Db 647 GluSerSerSerAlaProValProThrProSerSerSerThrThrGluSerSerSerAla 666  
Qy 1055 AGCGTGGCGAGAAAGACATTCGACAGCTGGATCC-----GTAAGAACGGCA 1102  
Db 667 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 686  
Qy 1103 AGAGTGTCTGTATGGCGATATGCTCTCTCGAAAGGCTTATAGGAAGGAGCA 1162  
Db 687 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 706  
Qy 1163 AGGCCAACCGTGAGATGACTTATTGACGAGACGCTCTTCGGTGGTACCGAGGTGCTTC 1222  
Db 707 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValPro----- 723  
Qy 1223 GTTTTGACACAGTTGGCAACCATTTGCGATGCTACAAATCTCGATGCTATCCCTCA 1282  
Db 724 ThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSerSerSer 743  
Qy 1283 AATCGTTGACGACAGTACAAAGACTACCTCCCTCGCTCGACCGTAAAGTGTGCTGCCG 1342  
Db 744 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 763  
Qy 1343 -----CATGCTCGATATTGTACCGCGGCTATCCCTGCGACACAGTCCCGGATA 1393  
Db 764 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThrProSer 783  
Qy 1394 TATTCAGAATGTATGC-----ACAAGAAATCAAGCGCAGCA 1435  
Db 784 SerSerThrThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGlu 803  
Qy 1436 AGAGTATG-----CAGACTTCGTATTCGACAGAGTGTGGTTCCTTATAGCG 1483  
Db 804 SerSerValAlaProValProThrProSerSerSerSerAla----- 818  
Qy 1484 ACAAGTTCATGCCATGCTCAAGTCCATGGACAGGAAAGTTTCCCAAGGCTATCGAGA 1543  
Db 819 ThrSerSerAlaProSerSerSerThrProPheSerSerThr----- 832  
Qy 1544 AAGATCCGGCAGTAGAGCTTCCCAAGAGCGTAAATAGTGTGCTCGCGCTATTACGGCGC 1603  
Db 833 -----GluSer 834  
Qy 1604 ATGCGATGGCCAAATGCCTATGCCATTGAGAAAGGCGAAGCGCTTTCTTTGCGGTTGTC 1663

Db 835 SerSerValProValProThrProSerSerSerThrThrGluSerSerSerAlaProVal 854  
Qy 1664 GTGAGATGTACCCCGACGTCTCTCGCGAGGATGCCAACT-----TCACCATGC 1714  
Db 855 SerSerSerThrThrGluSerSerValAlaProValProThrProSerSerSerAsn 874  
Qy 1715 GTATGAGTACGCGTCCATCAAGGGATATGAACCGCAGGACGGTGCCT----- 1762  
Db 875 IleThrSerSerAlaProSer-----SerIleProPheSerSerThr 888  
Qy 1763 -----GGTACAACATATACGACAGGCAAGGCGGTATTTGGAGAAGC 1804  
Db 889 ThrGluSerPheSerThrThrGluThrThrValThrProSerSerSerLysTyProGlySer 908  
Qy 1805 AGGATCTCAAGAGCGATGAGT-----TTCCCGTACAGGAGATA 1843  
Db 909 GlnThrGluThrSerValSerSerThrThrGluThrThrIleValProThrLysThrThr 928  
Qy 1844 TCCTCGACCTCTCCGACACCAAAATATGTC----- 1876  
Db 929 ThrSerValThrProSerThrThrThrIleThrThrValCysSerThrGlyThr 948  
Qy 1877 -----GCTATGCCGAGAACGGTCAAGTCCATATCGCTT 1909  
Db 949 AsnSerAlaGlyGluThrThrSerGlyCysSerProLysThrValThrThr----- 965  
Qy 1910 TCCTATCGAACACGACATCAGCGCGGTAACTCCGGTAGCCCGTATTCGATAGAACG 1969  
Db 966 ThrValProThrThrThrThrSerValThrSerSerSerThrThrThrThrThrThr 985  
Qy 1970 GCGGTCTGATCGGTCTTCTGATGCGCAACTGGGAAGCTATGAGTGGTGCATCGAGT 2029  
Db 986 ThrVal-----CysSerThrGlyThrAsnSerAlaGlyGluThrThrSerGly 1001  
Qy 2030 TCGAACCCG-----ATCTGCAGCGCACATACGCGTGGACATCC 2068  
Db 1002 CysSerProLysThrIleThrThrThrValProCysSerThrSer-----ProSer 1018  
Qy 2069 CTAAGTCTCTTCTATGATTGACAAATGGGTGTCAGTCCCGCTCATCC 2119  
Db 1019 GluThrAlaSerGluSerThrThrThrSerProThrThrProValThrThr 1035  
RESULT 11  
US-09-815-242-13382  
; Sequence 13382, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13382
; LENGTH: 883
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-815-242-13382

Alignment Scores:
Pred. No.: 0.0225 Length: 883
Score: 133.00 Matches: 104
Percent Similarity: 33.07% Conservative: 65
Best Local Similarity: 20.35% Mismatches: 166
Query Match: 3.48% Indels: 176
DB: 10 Gaps: 24

US-10-008-355-1 (1-2139) x US-09-815-242-13382 (1-883)

Qy 574 GATGATTC-----AAGGACGTTCTGATGTTGCTCCT 609
Db 26 AspValPheLysProSerGlyAspGlnLysAlaLysProTyrSerIleValIleProPro 45
Qy 610 CCCAGCTCTGTAGTAAGTTCGGAGGCGATACGGACAACCTGGATGTCGCCCGTCACACG 669
Db 46 ProAsnValThrGlyLysLeu-----HisLeu 54
Qy 670 GCGGACTTCACGGTATTCGCGGTGTATCCGTCGCGGACACCGCGCGCGGCAATACACG 729
Db 55 Gly-----HisAla 57
Qy 730 AAGGACATAAACCCCTAAGCCGTTTACTTCGTCGCGGTATCCATCGAAGGCTTACAAAG 789
Db 58 TrpAspThr-----ThrLeuGlnAspIleIleIleArgGlnLysArgMetGlnGlyPheAsp 76
Qy 790 GCTGACGACTATGCCATGACCATCGTTTCCGGGCGACTACGGATCGGTACTCCTC---ACT 846
Db 77 -----ThrLeuTrpLeuProGlyMetAspHisAlaGlyIleAlaThr 90
Qy 847 TCTTGGGTGTGAAGTCGTATCGAAGAACGAGAACATCTCGTATCGAAGTTCGCGGT 906
Db 91 GlnAlaLysValGluGluArgGlyGluGlyIleSerArgTyrAspLeuGlyArg 110
Qy 907 -----ATCAAGCAAGGATCTGGAAGGAGCAATCGAGCGACATCAGGCTACCC 954
Db 111 GluSerPheLeuThrLysValTrpGluTrpLysAsp-----GluTyrAlaThr 126
Qy 955 CGTATCAATATGCC---AGCAAGTATCTCAGAGTCTACTATTTGGAAGAAATTCGATC 1011
Db 127 ThrIleLysGluGlnTrpGlyLysMetGlyLeuSerValAspTyrSerArgGluArgPhe 146
Qy 1012 GGTATGAACCGGCTCGCTCGTCTTACGTATGATAGTGTACGTCGCGGAGGAAAGA 1071
Db 147 ThrLeuAspGluGlyLeuSer-----LysAlaValArgLys 158
Qy 1072 GCATTCGAGACTGTCGCTAAG-----AACGGCAAGAGTCTGCTCTATGCC 1119
Db 159 ValPheValAspLeuTyrLysLysGlyTrpIleTyrArgGlyGluPheIleIleAsnTrp 178
Qy 1120 GAT-----GTATTGCTCTCTCGAAAGGCTTATAAG----- 1152
Db 179 AspProAlaAlaArgThrAlaLeuSerAspIleGluValIleHisLysAspValGluGly 198
Qy 1153 -----GAAGGACCAAGGCCAACCGTCAGATGACT 1182
Db 199 AlaPheTyrHisMetAsnTyrMetLeuGluAspGlySerArgAlaLeuGluValAlaThr 218
Qy 1183 TATTTGAGCGAGACGCTCTCGGTGTCACCGAGGTGTTTCGTTTTCACAGTTTCCCAAC 1242
Db 219 ThrArgProGluThrMetPheGlyAspVal----- 228
Qy 1243 GCATTTGGGTACAAATTCCTGATGCTCATCGCGGTATTCCTCAATCGCTTTCACGACAAGTAC 1302

Db 229 AlaIleAlaValAsnProGlu-----AspProArgTyr 239
Qy 1303 AAGACTACCTCCCTCG-----CTCGACCGTAAGGTCTGCCGCCCATG 1347
Db 240 LysAspLeuIleGlyLysAsnValIleLeuProIleAlaAsnLysLeuIleProIleVal 259
Qy 1348 CTCGAT-----ATTGTACGCCGCGGT 1368
Db 260 GlyAspGluHisAlaAspProGluPheGlyThrGlyValValLysIleThrProAlaHis 279
Qy 1369 ATCCCTGCCGACAGCTCCCGATATATTCAAGATGTAATCGACAGAATAATCAAGGC 1428
Db 280 AspProAsnAspPheLeuValGlyGlnArgHisAsnLeuProGlnValAsnIleMetAsn 299
Qy 1429 GACACGAGAAGTAGTTCGACACTTCGTATTCGACAGAGCTGTGTTCCTTATAGCCACAAG 1488
Db 300 AspAspGlyThrMetAsnGluLeuValPheGluPheSerGlyMet-----AspArg 316
Qy 1489 TTCCATGCCATGCTCAAGTCCATCGACAAGAAAGTTTGCACAGGCTATCGAGAAAGAT 1548
Db 317 PheGluAlaArg-----LysAlaValAlaLysLeuGluGluIleGly 331
Qy 1549 CCGCAGTAGAGCTTTCACAGAGCGTAACTAGCTGCTGCTCGCGGT----- 1593
Db 332 AlaLeuValLysIleGluLysArgValHisSerValGlyHisSerGluArgThrGlyVal 351
Qy 1594 -----ATTCAGCCCGAT 1605
Db 352 ValValGluProArgLeuSerThrGlnTrpPheValLysMetAspGlnLeuAlaLysAsn 371
Qy 1606 GCGATGGCAATGCTATGCCATTCGAGAGGCAAGCGTCTTTCTTTTCCCGGTTTGGCT 1665
Db 372 AlaIleAlaAsnGlnAspThrGluAspLys-----Val 382
Qy 1666 GAGATGTACCCGGGACGTGCTGCGGAGC-----GAT 1698
Db 383 GluPheTyrProArgPheAsnAspThrPheLeuGlnTrpMetGluAsnValHisAsp 402
Qy 1699 GCCAACTTCACCATCGTATGACGTCACGCTCCATCAAGGGATATGAACCGCAGGACGT 1758
Db 403 TrpValIleSerArgGlnLeuTrpTrpGlyHis-----GlnIlePro 416
Qy 1759 GCCTGGTACAATATCATACGACAGGCAAGGCGCTATTGGAGAAGCAGGATCCTAAGAGC 1818
Db 417 AlaTrpTyrAsn-----AlaAspGlyLeuMetTyrValGlyGluAlaProGluGly 434
Qy 1819 GATGATTTGCGTCACGAGAGATATCTCGAC 1851
Db 435 AspGlyTrpThrGlnAspGluAspValLeuAsp 445

RESULT 12
US-09-815-242-13684
; Sequence 13684, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

574	QY	GATGTATTC-----AAGGAGTTCGTATGGTATTTGCTCTCT	609
26	Db	AspValPheLysProSerGlyAspGlnLysAlaLysProTyrSerIleValIlePro	45
610	QY	CCAGCTCTGTAGGTAAAGTTCGGAGCGCATACGGACAACCTGGATGTGCCGCGTCACACG	669
46	Db	ProAsnValThrGlyLysLeu-----HisLeu	54
670	QY	GCGAGCTTCAGCGTATTCGCGGTGTATGCCGTGCCGACAACCGCGCGCGAATACACG	729
55	Db	Gly-----HisAla	57
730	QY	AAGGACAATAAACCCCTAATAGCCGCTTACTTCGTCGCGTATCCATCAAGGCTACAAAG	789
58	Db	TrpAspThr---ThrLeuGlnAspIleIleLeaArgGlnLysArgMetGlnGlyPheAsp	76
790	QY	GCTGACGACTATGCCATGACCATCGGTTCCTCCGGCAGTACCGATCGCTACTCTC---ACT	846
77	Db	-----ThrLeuTrpLeuProGlyMetAspHisAlaGlyIleAlaThr	90
847	QY	TCATTGGGGTGGAGATCGTATCGAAGAACGAGACAACATCTCGTATCGAATTCGGCGGT	906
91	Db	GlnAlaLysValGluGluArgLeuArgGlyGluGlyIleThrArgTyrAspLeuGlyArg	110
907	QY	-----ATCAAGCAAGGCATCTGTGAAGGAAGCCATGAGCGGACATCAGCGTACC	954
111	Db	GluserPheLeuThrLysValTrpGluTrpLysAsp-----GluTyrAlaThr	126
955	QY	CGTATCAAAATATGCC---AGCAAGTATGCTCAGAGTGCTAACTATTTGGAAGAATTCGATC	1011
127	Db	ThrIleLysGluGlnTrpGlyLysMetGlyLeuSerValAspTyrSerArgLysPhe	146
1012	QY	GGTATCAACCGGGTCTCGCTCTTACGCTGATAGTCTGAACGCTGCCGAGGAAGA	1071
147	Db	ThrLeuAspGluGlyLeuSer-----LysAlaValArgLys	158
1072	QY	GCATTCCGACATGGATCGGTAAAG-----AACGGCAAGAGTGTCTGTATGGC	1119
159	Db	ValPheValAspLeuTyrLysGlyTrpIleTyrArgGlyGluPheIleAsnTrp	178
1120	QY	GAT-----GTATTGTCTTCTCTCGAAAAGGCTTATAAG-----	1152
179	Db	AspProAlaAlaArgThrAlaLeuSerAspIleGluValIleHisLysAspValGluGly	198
1153	QY	-----GAAGGAGCAAGGCCAACCGTGGATGACT	1182

; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 18  
; LENGTH: 26  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
; US-10-008-355-18

Alignment Scores:  
Pred. No.: 0.0237 Length: 26  
Score: 122.50 Matches: 25  
Percent Similarity: 96.15% Conservative: 0  
Best Local Similarity: 96.15% Mismatches: 0  
Query Match: 3.21% Indels: 1  
DB: 9 Gaps: 1

US-10-008-355-1 (1-2139) x US-10-008-355-18 (1-26)

QY 70 GCCGACAAAGGC---ATGTGCTCCTCAACGAACTCAATCAGGAGAAATCTGGATCGAATG 126  
|||||  
Db 1 AlaAspLysGlyMetMetTrrPLeuLeuAsnGlnLeuAsnGlnGluAsnLeuAspArgMet 20  
QY 127 CGTGAGCTCGGCTTACG 144  
|||||  
Db 21 ArgGluLeuGlyPheThr 26

## RESULT 14

US-10-192-584-6  
; Sequence 6, Application US/10192584  
; Publication No. US20030027987A1  
; GENERAL INFORMATION:

APPLICANT: TOKUNAGA, Eiji  
SAKAGUCHI, Masashi  
MATSUO, Kazuo  
HAMADA, Fukusaburo  
TOKIYOSHI, Sachio

TITLE OF INVENTION: NOVEL POLYPEPTIDE FROM HAEMOPHILUS

PARAGALLINARUM AND PROCESS FOR PREPARING THE SAME

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 624 Ninth Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/192,584

FILING DATE: 11-Jul-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/077,098

FILING DATE: 19-May-1998

APPLICATION NUMBER: PCT/JP97/03222

FILING DATE: 12-SEP-1997

APPLICATION NUMBER: JP 27,148/1996

FILING DATE: 19-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: KORNBAU, Anne M.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: TOKUNAGA-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-628-5197

TELEFAX: 202-737-3528

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2042 amino acids

TYPE: amino acid

STRANDEDNESS: single

; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-192-584-6

Alignment Scores:  
Pred. No.: 0.636 Length: 2042  
Score: 118.50 Matches: 140  
Percent Similarity: 36.03% Conservative: 87  
Best Local Similarity: 22.22% Mismatches: 223  
Query Match: 3.10% Indels: 180  
DB: 9 Gaps: 35

US-10-008-355-1 (1-2139) x US-10-192-584-6 (1-2042)

QY 550 AACCAATACCTTCCTCATCGTACGATGCTATTCAGAGGACGTCGTGTGTTGCTCTCT 609  
|||||  
Db 2 AsnLysValPheLysIleLysTyrSerValLysGlnGluMetIleVal----- 18  
QY 610 CCCAGCTCTGTAGTGTAGTTCGGAGCGGATACGGACAACCTGGATGTGCGCGGCACACG 669  
|||||  
Db 19 -----ValSerGluLeuAlaAsnAsnLysAspLysThrAlaSerGlnLysAsnThr 35  
QY 670 GCGAGCTTCAGCGTATTCCGC----- 690  
|||  
Db 36 HisAsnThrAlaPhePheGlnProLeuPheThrLysCysThrTyrLeuAlaLeuLeuIle 55  
QY 691 ---GTGTATGCGGTGCGGACACCGCGCCGCAATACAGCAGGACAATAAACCTTAT 747  
|||  
Db 56 AsnIleAlaLeuGlyAlaSerLeuPhePro-----GlnLeuAlaAsnAlaLysTrp 72  
QY 748 AAGCCCGTTACTTCGCTGCGGTATCCATGCAAGGCTACAAGGCTGACGAC----- 798  
|||||  
Db 73 LeuGluValTyrSerSerValLysLeuSerThrValSerAlaGlnSerAsnSerVal 92  
QY 799 -----TATGCCATGACCATC 813  
|||  
Db 93 AsnLeuAsnProSerGlyAlaGluSerValGlyThrAsnSerProGlnGlyValAlaIle 112  
QY 814 GGTTCCTCCGCGCAGTACGATCGGTACCTCACT-----TCTTGGGTGTGGAAGAT 864  
|||||  
Db 113 GlyTyrGlyAlaThrAsnAspArgSerAlaThrGlyAlaIleAlaLeuGlyVal----- 130  
QY 865 CGTATCGAAAACGAGACAATCCTCGTATCGAAGTTCGCGGTATCAAGCAAGCATCTGG 924  
|||  
Db 131 GlyValLysAsnGluThrLeuAlaLysAspSerIle---AlaIleGlyTyrGlyAlaLys 149  
QY 925 AAGGAAGCCATGAGCGCAGATCAGGCTACCGGTATCAATATGCCAAGTATGCTCAG 984  
|||  
Db 150 AsnGluSerThrAlaProSerSerValThrIleGlyLys-----Gln 163  
QY 985 AGTGCTAACTATTGGAAGAAATTCGATC-----GGTATGAACCGGGTCTCGCTGCTTT 1038  
|||  
Db 164 AlaIleAsnArgPheGluLysSerIleValMetGlyLeuAsn---AlaTyrThrGlnLeu 182  
QY 1039 GACGTGATAGGTGTAAGCGTCCGAGGAAAGAGCATTCGCGAGACTGGATCGTAAGAAC 1098  
|||  
Db 183 AspProArgGlyThrSerLys---GluThrArgGlnGlySerValIleGlyGluAsn 201  
QY 1099 GGCAAGAGTGTCTTATGCGGATGCTATGTCTTCTCTCGAAAGGCTTATTAAGGAAGA 1158  
|||  
Db 202 AlaLysSerAla-----GlyAsn-----GlnSerValSerLeuGlyGlnAsnSerTrp 217  
QY 1159 GCCAAGGCCAACCGTCGATGACATTTATTGACGAGAGCGCTTCCTGGTGGTACCGAGGTG 1218  
|||  
Db 218 SerLysThrAsn-----SerIleSerIleGlyAlaGlyThr----- 229  
QY 1219 GTTCGTTTTCACAGCTTTGCCAACGATTTGGCTACAAATCCTGTGCTCATGCCGTATC 1278  
|||  
Db 230 -----PheAlaGluGlyLysSerSerIleAlaIleGlyThrAspLysIleSerGly--- 246  
QY 1279 CTCAAATCGCTTGACGACAAGTACAAAGACTACCTCCCTCCCTCGCTCAGCGTAAAGTGCTG 1338  
|||||

```
Db 247 -----ThrLysTyrAsnAspLysLeuProAlaThrAlaTrpAsnGlyThr 261
Qy 1339 CCCGCATGCTCGATATTGTACGCGCGGTATCCCTCCGCGCAAGCTCCCGCATATATTC 1398
Db 262 GlyThr-----ValProLysAsnSerIleTrpAspIlePhe 273
Qy 1399 AAGATGTA---ATCGACAAGAAATTCAAAGC-----GACACGAAG----- 1437
Db 274 SerGluLeuTyrMetGlyLysGlnThrAsnGlyArgAspTyrAspThrThrArgAsp 293
Qy 1438 -----AAGTATGCAGACTTC-----GTATTCAC 1461
Db 294 ProAsnLysProGluAlaPheTyrLysPheSerAspPheLysGlyLysTyrValAsnThr 313
Qy 1462 AAGAGTGTGGTTCCT---TATAGCCACAAGTTCATCCATGCCTCAAGTCCATGGACAAG 1518
Db 314 ProThrAlaSerProThrTyrAlaGlyLysLeuGlyAlaIle----- 327
Qy 1519 GAAAGATTGGCAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAAGAGCGTA--- 1575
Db 328 -----AlaLeuGlySerArgThrIleAlaAlaGlyGluMetSerThrAlaValGly 344
Qy 1576 ---ATAGCTGTGCTCGCGCTATTTCAGGCCGATCGCATGGCCATGCCATGCCATTGAG 1632
Db 345 SerLeuAlaPheAlaLeuAlaAspArgSerThrAlaMet----- 357
Qy 1633 AAGGCAAGCGCTTTCTTTGTCGGGTTTGGGTGAGAGTATACCCGGAGCTGCTCGCGC 1692
Db 358 ---GlyLeuArgSerPheValAla-----LysAspAlaValGlyThrAlaIleGly 374
Qy 1693 AGCGATGCC-----AACTTCACCATCGGTATG 1719
Db 375 GluGluSerArgThrPheAlaLysAspSerValAlaIleGlyAsnLysThrGluAlaSer 394
Qy 1720 AGCTACGCTCCATCAAG----- 1737
Db 395 AsnAlaGlySerMetAlaTyrGlyTyrLysAlaLysAlaValGlyAlaIleAla 414
Qy 1738 ---GGATATGAACCGACGAGCGGTGCTTACAACTATCATACACAGCAAGGCGCTA 1794
Db 415 IleGlyThrGluValAlaAlaGlyAlaLysPheAsnSerHisGlnThrGlyAsnLeuLeu 434
Qy 1795 TTGGAG-----AAGCAGGATCCCTAAGACGAGTGGTGTTCGCCGTA 1833
Db 435 GlnAspAsnAlaTyrAlaThrLeuLysAsnAlaAspLysSerAsp---ThrLys 453
Qy 1834 CAGGAGATATCTCGACCTCTTCGCGCACCAAACTATGTCGCTATGCCGAGAACGCT 1893
Db 454 ThrGlyAsnAlaIleThrValPhe---ThrGlnSerPheAspAsnMetLeuThrAsnGly 472
Qy 1894 CAGCTCCATATCGCTTTCTCTATCGAACACGACATC----- 1929
Db 473 -----LeuProLeuValSerGluAsnGluThrTyrLeuThrThrSerAlaGlyAla 489
Qy 1930 -----ACGGCGGTAACTCCGGTAGCCCGGTATTCGATAGAAGCGCGTGTGATC 1980
Db 490 IleLysLysThrAlaThrAspSerSerAlaGlyGlyLysAsnAlaIleAlaIle 509
Qy 1981 GGTCTTGTGTTGCGTGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCCGACCCGAT 2040
Db 510 GlySerLys---ThrPheAlaSer---LysAlaAsnSerValAlaLeuGlySerTyrAlaLe 528
Qy 2041 CTGCAGCGCACAATCAGCGTGGACATCCGCTACGTTCTCT-----TCATGATTGAC 2091
Db 528 uAlaAspAlaGlnAsnAlaPheAlaLeuGlySerTyrSerPheValGluSerSerAlaTh 548
Qy 2092 AAATGGGTGACGTCCCGCGCTCATCC 2119
Db 548 rAsnThrIleThrIleGlyValGlySer 557
RESULT 15
US-09-920-552-103
; Sequence 103, Application US/09920552
```

```
; Patent No. US20020094576A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, David J.
; APPLICANT: Weiss, Robin A.
; APPLICANT: Venables, Patrick
; TITLE OF INVENTION: Material and Methods Relating to a No. US20020094576A1el Retro
; FILE REFERENCE: Abbott Labs
; CURRENT APPLICATION NUMBER: US/09/920,552
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 09/280,329
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: GB 9806649.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/115,288
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 103
; LENGTH: 1814
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-920-552-103

Alignment Scores:
Pred. No.: 0.656 Length: 1814
Score: 118.00 Matches: 162
Percent Similarity: 33.29% Conservative: 89
Best Local Similarity: 21.49% Mismatches: 252
Query Match: 3.09% Indels: 251
DB: 10 Gaps: 43
```

US-10-008-355-1 (1-2139) x US-09-920-552-103 (1-1814)

```
Qy 255 TACCAACCAACACCTCCGATACGGTGCTATCCAGAGCCAAAGCGGTGGATCAGACTA 314
Db 47 TyrSerProProArgSerPheSerLysArgGlyLysArgGlyArgLysIle 66
Qy 315 TCTGCG-----CGATGGTTT-----CGTTTCTCGCAC 341
Db 67 HisAlaPheValArgAlaValGluMetSerProTrpPheLeuGlnTrp-ArgG1 86
Qy 342 GATGGTGAGGAGCTTCGATTCGGGTCTTCCGTGAAGTATCTGCCAGAGATCGTAA 401
Db 86 nValGlyTrpLysArgGluAlaGlySerIleGluValLeu---LysLeuTrpG1 105
Qy 402 GGTACGCAAGGTAGAGGACAGCTCAAGGTATCACTGACGAGATGGAG----- 453
Db 105 nLeuValAspGlnVal-----LeuGlnAspPheGluGlyIleG1 119
Qy 454 ---CGTCTGCGCAAGCTCAGGAGGTATGCCAAGTGGCCAAAGAAATGCAGA 509
Db 119 uLeuLeuLeuAsnGluGlySerGluLeuGlnGluLeu-----TrpGlnGlySerG1 137
Qy 510 CGAGAACCACTCTGCATCTAGAGCCTTTCTATTCACACAGAACTCTTCCTCATCT 569
Db 137 nGluSerSerTyrCys---AsnGluGlnValArgArgLysSerLysAspLeuGlu 156
Qy 570 CTACGATCATTC-----AAGGAGCTTCGTATGTTGCTCTCTCCAGCTCTGT 620
Db 156 rGlnGluAlaTyrValGlyGluLysGluAlaGluGlyThrSerProProArgGluMe 176
Qy 621 AGGTAAGTTCGAGCGCATCGGACAACTGGATGTGGCGCGTCAACACGGCGACTTCAG 680
Db 176 tPro----- 177
Qy 681 CGTATTCGCGTGTATCCGGTGGGACACACGGCGGCCCAANTACACAGGACAATAA 740
Db 178 -----ArgIleTyrProGlyProThrArgLeuAlaAsnSerThrIleLysAspGlu 195
Qy 741 ACCC-----TATAAGCCCGTTTAC-----TTCCGCTGCCGTATCCATCAAGGCTA 785
Db 195 rProProArgTyrArgProLeuValLeuProLysPheProValLysSerProPh 215
```

```
QY 786 CAAGGCTGACGACTATGCCATGACCATCGGTTCCTCCGGGCGAGTACGATCGCTACCTCAC 845
Db 215 e-----PheProHisGluIleGlu----- 222
QY 846 TTCTTGG-----GGTGTGGAAGCATCGTATCGAATA 875
Db 223 -GluTrpGluGluGluGlyGlyValMetGluGlyValGlyMetGluGluIleGlu-- 241
QY 876 CGAACAATCTCTCGATCGAA-----GTTCCGGGT-----ATCAAGCAAGCGCATCTG 923
Db 242 -----ArgProArgThrGluSerLysGlnArgGlyAlaValTyrArgLysProAlaTr 259
QY 924 CAAGGAAGCCATGAGCCGACATCAGGCTACCCGTATC----- 960
Db 259 pGluGlnTyrGlnGluProGlyGlnValPheProValIleGlnAspGlnGlnGluAsnAr 279
QY 961 -----AAATATGCCAGCAAGTATGCTCAGAGTCTAACTATTG 998
Db 279 gGlyTrpThrProLeuAspHisLysLeuLysGluLeuGlnSerValGlnLeuTy 299
QY 999 GAAGAATTCGATCGGTATGAACCGCGTCTCGTCTTTGAGCTGATAGTGTGTAAGCG 1058
Db 299 rGlyProHisAlaSerTyrThrGlnAlaIle-----LeuAspAsnIleGlyGlnGlnG 317
QY 1059 TCGGAGGAAGAGCATTCGACACTCGATCGGTAGAACGCGCAAGAGTGTCTGTAT-- 1116
Db 317 yLeuIleProGlu-----AspTrp-----ArgAsnLeuValLysAlaValLeuG 332
QY 1117 ----GGGATGATTTGCTCTCGAAGAGCTTATAAGGAAGGAGCGCAAGCCCAACCG 1172
Db 332 yGlyGlyAspPheLeuLeuTrpLeu---AlaAlaTyrLysGluPheAlaArgGluTyrSe 351
QY 1173 TGAGATGACTATTGTGAGGAGAGCGTCTCTCGGTGTACCGAGTGGT----- 1221
Db 351 rLysGlnAsnTyrGlnAlaGlyAsnGlnAlaTrpAspGluGluMetLeuIleGlyGluG 371
QY 1222 -CGTTTTCACAGTTTCCCAACGATGGCTACAAATCTCGATGCTCATGCCGATPCT 1280
Db 371 yArgPheAla-----AsnProAspGluGlnAla----- 380
QY 1281 CAATCGCTTGACGACAAGTACAAAGACTACCTCCCTCGCTCGACCGTAAGGTGTCGC 1340
Db 381 -----ArgLeuProThrAlaValLeuLe 388
QY 1341 CGCATGCTCGATATTGTACGCGCGGTATCTCCTCGCGACAAGCTCCCGCATATATTC 1400
Db 388 uGlnValArgGluIleAlaArgAlaTrpLysValIleProArgLysGlyGluIleAr 408
QY 1401 GAATGTAATCGACAAGAAATTCAAAGCGGACACGAGAAGTATGCAGATTCGTA----- 1455
Db 408 gHisSerLeuThrLysIleValGlnGlyProThrGluProTyrAlaAspPheValAsnAr 428
QY 1456 -----TTCGAC-----AAGAGTGTGCTCTCT----- 1476
Db 428 gLeuMetGluAlaAlaGlyAsnIlePheGluThrValGluGluAlaMetProLeuValAr 448
QY 1477 -----TATAGCACAAGTTCATGCTGCTCAAGT-----CCATGGACAA 1517
Db 448 gArgLeuAlaTyrGluGln-AlaAsnLysThrCysArgGluAlaLeuArgProTrpGlnH 468
QY 1518 GGAAGCTTTCCCAAGGCTATCGAGAAGATCCGGCAG-----TAGAGCT 1562
Db 468 isLysAspIleProThrPheLeuLysIleCysArgAspValMetAspValAlaSerG 488
QY 1563 TTCCAAGAGCGTAATAGTCTGCTCGCGTATTTCAGCGCGATCGCATGCCAATGCCCTA 1622
Db 488 lySerHisAla-----AlaValAlaValThrArgGlnPheArgGlnAsnMet---- 503
QY 1623 TGCCATTGCAAGGCGAAGGCTCTTTCTTTTTCGCGTTTCGTCGAGATGATCCCGCGACG 1682
Db 504 -----Arg-GlyArgIleCysPheArgCysGlyArgGluGlyHisLeuMetArgGlu 520
QY 1683 TGCTCTCCGAGCGATGCCAACTTCACCATGCGGTATGAGCTACGGCTCCATCAAGGGATA 1742
```

Search completed: May 16, 2003, 13:03:11

Job time : 124.5 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:22:51 ; Search time 79.5 Seconds  
(without alignments)  
5173.120 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 3820

Sequence: 1 atgcaaatgaaattaaaag.....aagagctgaattgatctaa 2139

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool/US10008355/runat\_16052003\_110338\_9364/app\_query.fasta\_1.2311  
-DB=pir\_73 -QPMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45  
-LOCALALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355.ecgn\_1.1.111.0runat\_16052003\_110338\_9364 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

pir\_73: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	907.5	23.8	716	2 G82627	hypothetical prote
2	156	4.1	1791	2 T02345	hypothetical prote
3	154.5	4.0	1609	2 S25345	probable membrane
4	149	3.9	528	2 I47141	gastric mucin (clo
5	145	3.8	1367	1 S48478	glucan 1,4-alpha-g
6	143.5	3.8	776	2 S44784	C30C11.4 protein -
7	141	3.7	948	2 T11678	hypothetical prote
8	139.5	3.7	1802	2 S69703	HKR1 protein precu
9	139	3.6	4543	1 A53102	alpha-2-macroglobu
10	138	3.6	477	2 S53362	mucin 5AC (clone J
11	137	3.6	3570	2 T45025	mucin MUC5B, trach
12	133.5	3.5	1104	2 A60999	alpha-amylose (EC
13	133	3.5	883	2 D97933	valine-tRNA ligase
14	132	3.3	571	2 C71528	probable pts pep p

c

ALIGNMENTS

RESULT 1

G82627  
hypothetical protein XF1887 [imported] - Xylella fastidiosa (strain 9a5c)  
C/Species: Xylella fastidiosa  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C/Accession: G82627  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A/Reference number: A82515; MUID:20365717; PMID:10910347  
A/Note: for a complete list of authors see reference number A59328 below  
A/Accession: G82627  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-716 <STM>  
A/Cross-references: GB:AE004008; GB:AE003849; NID:99106961; PIDN:AAF84693.1; GSPDB:GN  
A/Experimental source: strain 9a5c  
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Krieger, J.E.; Kuramae, E.E.; La  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Marinho, C.L.; Marques, M.V.; Martins  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Menezes, C.F.M.; Miracca, E.C.; Miyaki, C.  
A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
Rodrigues, V.; Rosa, A.J. de M.; da Silva, F.R.; da Silva, A.C.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A/Reference number: A59328  
A/Contents: annotation  
C/Genetics:  
A/Gene: XF1887

Alignment Scores:  
Pred. No.: 7.33e-54  
Score: 907.50  
Percent similarity: 49.21%  
Length: 716  
Matches: 239  
Conservative: 135

Best Local Similarity: 31.45% Mismatches: 291  
Query Match: 23.76% Indels: 95  
DB: 2 Gaps: 21

US-10-008-355-1 (1-2139) x G82627 (1-716)

```
Qy 1 ATGCAATGAAATTAATAAGATTTCTTCGCGAGCAGCCCTGCTGTTGGTGCTTCAGGG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MetArgPheAsnLeuSerLeuSerValLeuAlaThrLeuIleThrValAspSer--- 19
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 61 GTAGCCAAAGCCGCAAGGATGGTCTCTCAACGAACTCAATCAGGAGAACTGGAT 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20 ---ThrHisAlaGlyGluGlyMetTrpValProGlnGlnLeu---ProGluIleAlaGly 37
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 121 CGAATGCGTGAGCTCGGCTTACGCTCCGTTGGATTCCGCTACAGTTTCGACAAGCCG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38 ProLeuLysGlnAlaGlyLeuGlnLeuSerProGluGlnLeuSerAsnLeuThrGlyAsp 57
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 181 TCCATTGCGCAATGCGGTGTTATCTTCGTCGCGGATGTACCGGTATCACAGTGTCCGAT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 ProMetGlyAlaValAlaValSerLeu-----GlyAsnCysThrAlaSerLeuValSerPro 75
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 241 CAGGCGCTGATCTTACCAACACACACTCGGATACGCTGCTATCCAGAGCCAAAGCAGC 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 GluGlyLeuValIleThrAsnHisCysAlaTyrGlyAlaIleGlnLeuAsnSerThr 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 301 GTGGATCAGACTATCTCGCGATGGTTTCTTCGTCGCGATGGGTGAGGAGCTCCG 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 ProLysLysAsnLeuIleLysGluGlyPheAsnAlaLeuThrGlnAlaAspLysValSer 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 361 ATT---CCGGTCTTTCGGAAGTATCTGCGCAAGATCGTGAAGTAAACGACAAAGTA 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 116 AlaGlyProAsnAlaArgIleTyrValLeuGluGlnIleThrAspValThrAlaGlnAla 135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 418 GAAGCAGACTCAGGGTATCAGTCAGGATGAGGAGCTCTGCGCAAGCTCAGAGAGTA 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 LysAlaLeuAlaAlaAlaGlyAsnAspProPheLysArgThrThrAlaLeuGluThr 155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 478 TGC-----CAAGACTCGCCAAAAGAAATGACAGACGAGAACCACTCTGCATCGTA 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 PheSerLysGlnGluIleAlaLysCysGluGlu---GluGlnGlyTyrArgCysGlnPhe 174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 532 GAGCCTTTCTATTCACCAACAGAAATACCTCTCATCTGCTACGATGATTCAGGACGTT 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 175 PheSerPheAlaGlyGlyAsnThrTyrArgValPheLysAsnLeuGluIleLysAspVal 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 592 CGTATGGTATTTCCTCCAGCTCTAGTAAAGTTCGAGCGGATACGACAACTGG 651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 ArgLeuValTyrAlaProGlnGlySerValGlyLysPheGlyLysValAspAsnTrp 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 652 ATGTGCGCGCTCACACGGCGACTTACGCGTATTCCGCGTGTATCCGCGTCCGACAAAC 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 MetTrpProArgHisThrGlyAspPheSerPheTyrArgAlaTyrValGlyLysAspGly 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 712 CGCGCGCGCAATACAGCAAGCAATAAACCTATAAGCCGTTTACTTCGCTGCCGTA 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 LysProAlaSerPheSerLysGluAsnIleProTyrArgProLysHisTrpLeuLysPhe 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 772 TCCATGCAAGGCTACAGGCTGACGACTATGCCATGCCATCCGCTTCCCGGGCAGTACG 831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 SerAspGlnProLeuLysGlyAspPheValMetValAlaGlyTyrProGluIleArgThr 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 832 GATCGCTACCTCCTCTTGGGTGTGGAAGATCGTATCGAAAACGAGAACAACTCCCTCGT 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 AsnArgTyrAlaLeuVal-----AlaGluPheGluAsnThrAlaHis 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 892 ATCGAAGTTCGCGGTATCAACCAA-----GGCATCTGGAAGGAAGCC 933
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 TrpThrTyrProValIleGlyGlnHisPheLysAsnLeuIleAlaLeuIleGluAlaAla 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 934 ATGAGCGCAGATCAGGCTACCGGTATCAATATGCCAGCAGTATGCTCAGAGTGTCTAAC 993
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 SerLysGlnAsnProAspIleGlnValLysTyrAlaSerThrLeuAlaGlyLeuAsnAsn 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 994 TATTGGAAGAATTCGATCGGTATGAACCGCGTCTCGCTCTTACGCTGATAGGTGCT 1053
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 ThrSerLysAsnPheAspGlyGlnLeuAspGlyPheArgArgIleAsnAlaIleGlyGln 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1054 AAGCGTCCGAGGAAAGAGCATTCGACACTGGATCCGTAAGAAGCGC----- 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 LysGlnSerGluGluThrAlaValLeuAlaThrLeuLysGlnGlnGlyIleArgGlyHis 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1102 AAGAGTGTCTATTCGCGATGTTCTCTCTCGAAAAGGCTTATAAGGAAGGACCC 1161
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 GluAlaLeuAlaAlaHisGlnThrLeuValAspLeuThrGluGlnTyr----- 384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1162 AAGGCCAACCCGTGAGATGACTTATTGAGCAGAGCGCTCTTCGCTGCTACCGAGTGT 1221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 LysAlaAsnGlnAspArgAspPheValLeuGlyGln---PheAsnGlySerGlyValIle 403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1222 CGTTTTGACAGTTTGGCAACGCATTTGGCT-----ACAAATCCTGATGCTCAT--- 1269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 GlyValAlaValAsnLeuTyrArgLeuAlaIleGluArgThrLysSerAspAlaGlnArg 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1270 -----GCCGGTATCTCAATCGCTTGAC 1293
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 GluAlaGlyTyrGlnGluArgAspLeuProThrIleGluGlyAsnLeuLysGlnMetGlu 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1294 GACAAGTACAAAGACTACCTCCCTCGCTCGACCGCTAAGGTG----- 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 -----ArgArgTyrLeuProGluMetAspArgGlnMetGlnGlnTyrTrpLeuThr 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1336 -----CTGCCCGCCATCTCGATATTGTACCGCGCGGTATCCCTGCC----- 1377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 GluTyrAsnLysLeuPro-----ValLysGlnArgValAlaAlaIleAsp 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1378 -----GACAGCTCCCGATATATTCAAGATGTAATCGCACAAATTCAAA 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 ValTrpLeuGlyAspGlyIleProAlaThrLeuLysArgLeuGlyAspThrLysLeuSer 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1426 GCGCACACGAAGATGACGACTTCGTAATCGACAAAGAGTGTGTTCTTATAGCGAC 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 496 SerSerGluGluArg----- 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1486 AAGTTTCATGCCATCTCAAGTCCATGGACAAAGAAAGTTTGCAGAGCTATCGAGAA 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 -----LeuLysTrpPheAsnAlaAspArgAlaAlaPheGluSerSerGln 515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1546 GATCCGCGCAGTAGACTTTCCAAAGCGCTAATAGCTGCTCGCTATTCAGGCCGAT 1605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 AspProAlaIleArgTyrAlaValAlaIleMetProAlaLeuLeuGluIleGluArgGln 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1606 GCGATGGCCAATGCTATGCCATTGAGAAAGGCAAGCGCTCTTTCTTTCGCGTTGCGT 1665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 AsnLysIleArgThrGlyGluLeuLysAlaArgProIleTyrLeuGlnAlaLeuAla 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1666 GAGATGTACCC-----GGACGTGCTCTGCCGAGCGATGCCAATCCATCCGCTATG 1719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 AspTyrAsnLysSerHisGlyLysPheValTyrProAspAlaAsnSerSerLeuArgIle 575
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1720 AGCTACGCTCCATCAAGGATATCAACCGCAGCGGCTGCTGCTACACATCATACG 1779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 ThrPheGlyHisValLysGlyTyrSerProLysAspGlyValGluTyrThrProPheThr 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1780 ACAGCGAAGGCGTATTGGAAAGCAGGATCCTAAGACGATAGTTTCCGCTACAGAG 1839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 ThrLeuGlnGlyValMetAlaLys---AsnThrGlyValGluProPheAspSerProLys 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1840 AATATCTCGACCTCTTCCGCCACCAAAACTATGCTGCTATGCCGAGAAC-----GGT 1893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 SerLeuIleAsnAlaIleLysAlaLysSerTyrAlaAsnLeuAlaAspGlnArgIleGly 634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1894 CAGCTCCATATCGCTTCTTCATCAACACGACATCAGCGCGGTAACTCCGGTAGCCCC 1953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 ThrValProValAsnPheLeuSerAspLeuAspIleThrGlyLysSerGlySerPro 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```





```
Db 972 GluValCysSerAsnSerIleCysThrProThrValIleThrSerValGlnMetArgSer 991
QY 881 -----ACAATCCCTCGTATCGAAGTTCGGGTATCAAGCAAGGCTCTGGAAGAAG 931
Db 992 ThrProPheProTyrLeuThrSerSerThrSerSerSerLeuAlaSerThrLysLys 1011
QY 932 CCATGACGCGAGATCAGGTCACGGTATCAAAATATGACCAAGATGCTCAGAGTGCTA 991
Db 1012 -----SerSerLeuGluAlaSer----- 1017
QY 992 ACTATTGGAGAATTCGATCGGTATGAACCGCGGTCTCGCTCTGACGTGATAGTGC 1051
Db 1017 ----- 1017
QY 1052 GTAAGCGTCCGAGAAAGAGCATTCGCAGACTGGATCCGTAAGAACGGCTTAAAGAGGCAAGGCCAAC 1171
Db 1018 -----SerGluMetSerThrPheSerValSerThr---GlnSerLeu 1030
QY 1112 TCTATGCGCATGATTCTCTCTCTCTCTGAAAGGCTTATAAGAGGAGGCAAGGCCAAC 1171
Db 1031 ProLeuAlaPheThrCys-----SerGluLysArgSerThr 1042
QY 1172 GTGAGATGACTTATTGAGCGAGAGCGCTCTCGGTGGTACCGAGGTGGTTCGTTTCAC 1231
Db 1043 -----ThrSerValSerGlnTrp----- 1048
QY 1232 AGTTTGCAACGCGATTGGCTACAAATCCTGATGCTCATCGCGGTATCCTCAAAATCGCTTG 1291
Db 1049 -----SerAsnThrValLeuThrAsnThrIleMetSerSerSerAsnValIle 1065
QY 1292 ACGACAAGTACAAAG-----ACTACCTCCCTCGCTCGACCGCTAAGGTGCTGCCCG 1342
Db 1066 SerThrAsnGluLysProSerSerThrSerProTyrAsn----- 1079
QY 1343 CCATGCTCGATATTGACGCGCGGTATCCTCGCGACAGACTCCCGATATATTCAGA 1402
Db 1080 -----PheSerSerGlyTyrSerLeuProSerSerSerThrProSerGlnTyr 1095
QY 1403 ATGTAATCGACAAGAAATTCAAAGGCGACAGCAAGAAAGTATGACAGACTTCGATTTCGACA 1462
Db 1096 SerLeuSerThr-----AlaThrThrIleAsnGlyIleLysThrValTyrThrThr 1113
QY 1463 AGAGTGTGGTTCCTATTAGCGCAAGTTCATGCCATCCATGCTCAAGTCCATCCAGCAAGAAA 1522
Db 1114 -----TrpCysProLeuAlaGluLysSerThrValAlaAlaSerSerGlnSerSerArg 1131
QY 1523 AGTTTGCAAGGCTA-----TCGAGAAAGATCCGCGAGTAGAGCTTCCAAAGACGCTAA 1576
Db 1132 SerValAspArgPheValSerSerSerLys--ProSerSerSerSerLeuSerGlnThrSerI 1151
QY 1577 TAGCTGCTGCTCGGCTATTTCAGCGCCGATGCGATGGCAATGCCATATGCCATTGAGAAGG 1636
Db 1151 le-----GlnTyrThrLeuSerThrA 1158
QY 1637 GCAAGCGCTTTTCTTTCGCGGTTTTCGCTGAGATGTACCC----- 1677
Db 1158 laThrThrThr---IleSerGlyLeuLysThrValTyrThrTrpCysProLeuThrS 1177
QY 1678 -----GGAGGTGCTCTCGGACCGATGCCAAGTTCACCTTCCATCGGTATGAGCT 1723
Db 1177 erLysSerThrLeuGlyAlaThrThrGlnThrSerSerThrAlaLysValArgIleThr- 1196
QY 1724 ACGGCTCCATCAAGGGATATGAACCGCAGGAGCGGTGCTGTGTAACACTATCATACGACAG 1783
Db 1197 SerAlaSerSerAlaThrSerThrSerIleSerLeuSerThrSerThrGluSerGluSer 1216
QY 1784 GCAAGGCGGTATTCGGAAGCAGGATCCTAAGACGCGATGAGTTTCGCGTACAGGAGAATA 1843
Db 1217 SerSerGlyTyrLeuSerLysGlyValCysSerGlyThrGluCysThrGlnAspValpro 1236
QY 1844 TCCTCGACCTCTCCCGACCAAAACATATGGTCTGCTATGGCGAAGACGGTCAGCTCCATA 1903
Db 1903 ----- 1903
```

```
Db 1237 ThrGlnSerSerSerProAlaSerThrLeuAlaTyrSerProSerValSerThrSerSer 1256
QY 1904 TCGCTTTTCTTATCAACAACAGCATCACGGCGGTAAC-----CCG 1945
Db 1257 SerSerSerPheSerThrThrThrAlaSerThrLeuThrSerThrHisThrSerValpro 1276
QY 1946 GTAGCCCGGATTTCGATAGAAGCGCGCTGATCGGTCTTGTCTTCGATGGCAACTGGG 2005
Db 1277 LeuLeuProSerSerSer-----SerIleSerAlaSerSerProSerSer 1291
QY 2006 AAGCTATGAGTGGTGCATCGATCGATTGCAACCCGATCTGCAGCGGCACAAATCAGCGTGGACA 2065
Db 1292 ThrSerLeuLeuSerThrSerLeuProSerProAlaPheThrSerThrLeuProThr 1311
QY 2066 TCCCTACGTTCTTTCATGATTCACAAATGCGGTGACAGTCCCGCTCTCATCCAAAGAGC 2125
Db 1312 AlaThrAlaValSerSerSerSerThrPheIleAlaSerSerLeuProLeuSerSerLysSer 1331
RESULT 4
I47141
gastric mucin (clone PGM-2A) - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000
C:Accession: I47141; S55315
R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Gastroenterology 106, 200, 1994
A:Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a nov
A:Reference number: I47141; MUID:94102478; PMID:7506218
A:Accession: I47141
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-528 <TUT>
A:Cross-references: EMBL:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Biochem. J. 308, 89-96, 1995
A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
A:Reference number: S55315; MUID:95275264; PMID:7755593
A:Accession: S55315
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-528 <TUT>
A:Cross-references: GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
C:Superfamily: pig submaxillary mucin
Alignment Scores:
Pred. No.: 0.0419 Length: 528
Score: 149.00 Matches: 154
Percent Similarity: 34.22% Conservative: 91
Best Local Similarity: 21.51% Mismatches: 243
Query Match: 3.90% Indels: 228
DB: Gaps: 28
US-10-008-355-1 (1-2139) x I47141 (1-528)
QY 68 AAGCCGACAAGGATGTGGCTCTCTCAACGAACATCAATCAGGAGAACTTGTGATCGAATGC 127
Db 5 GlnProSerSerSerSerSerSerProThrThrSer----- 16
QY 128 GTGAGCTCGGCTTTACGCTCCGTTGGATTGCTCTACAGTTTCGACAAAGCGTCCATTG 187
Db 17 ---ThrThrSerValGlnSer-----SerSerSerSerSerVal 28
QY 188 CCAATGCGGTGTTATCTTCGTTGGTGGCGGATGACCGGTATCACAGTGTCCGATCAGGCGC 247
Db 29 ProIleProSerThrThrSerValGln-----ProSerSerSer 41
QY 248 TGATCTTTACCAACACACATCGCGATACGTTGCTATCCAGAGCCAAAGACAGCGTGGATC 307
Db 42 GlySerAlaProThrThrSerAlaThrSerVal----- 52
QY 308 ACGACTATCTCCGCGATGGTTTCGTTCTCCGACGATGGGTGAGGAGCTTCGATTCGG 367
Db 52 ----- 52
```

Qy 368 GTCTTTCCGTGAAGTATCTCGCAAGATCGTGAAGTAAACGACAAAGGTAGAAGGACAGC 427  
Db 53 -----GlnThrSerSerSerSerSer 59  
Qy 428 TCAGGGTATCACTGACGAGATGAGCGTCTGCGCAAAAGCTCAGGAGGTATGCCAAGAAC 487  
Db 60 ProProIleSerThr----- 65  
Qy 488 TGGCAAAAAGAAAGTGCACAGACAGAACCAACTCTGCATCGTAGAGCCTTTCTATTCCA 547  
Db 66 -----IleSerValGlnThrSerSerSer-----SerValPro 77  
Qy 548 ACAAGAACTACTCTCATCTCTACGATGATTCAAGGACCTTCGTATGTTGCTC 607  
Db 78 ThrThrSerThrThrSer-----Val 84  
Qy 608 CTCGCCAGTCTGTAGTAAGTTCCGAGGCGATACGGACAACTGGATGTGCGCGTCACA 667  
Db 85 GlnProSer-----SerSerSerSerAlaProThr-----ArgAlaThr 98  
Qy 668 CGGCGACTTCAGCGTATTCGCGGTGATGCCG----- 700  
Db 99 SerValGlnSerSerSerSerSerSerAlaProIleSerSerThrThrSerValGlnPro 118  
Qy 701 -----GTGCCGACAAACCGCGCGCGAATACAGCAAGGACAAATAAACCCCT 745  
Db 119 SerSerSerGlySerValProThrThrSerAlaThrSerValGlnSerSerSerSer 138  
Qy 746 ATAAGCCGTTTACTTCGTCGCCGTATCCATGCAAGGCTACAAGGCTGACGACTATGCCA 805  
Db 139 SerAlaProThrThrSer-----AlaThrSerValGlnProSerSerSerSer 154  
Qy 806 TGACCATCGGTTTCCCGGCGAGTACGATCGCTACCTACCTCTCTTGGGGTGTGAAGATC 865  
Db 155 SerProProIleSerSerThrValSerValGlnProSer----- 167  
Qy 866 GTATCGAAACGAGAACAACTCCTCGTATCGAAGTTCCGCGTATCAAGCAAGGCATCTGGA 925  
Db 168 ---SerSerSerSerAlaProThrThrSerAlaThrSerValGlnProSerSerSer 186  
Qy 926 AGGAAGCATGAGCGAGATCAGCTACCGTATCAATATGCAAGCAAGTATGCTCAGA 985  
Db 187 SerProPro-----IleSerSerThrValSerValGlnThrSerSerSerSer 203  
Qy 986 GTGCTAACTATTGGAAGAAATTCGATCGGTATGAACCGGGTCTCGCTGCTTGCACGTGA 1045  
Db 204 ValProThr-----ThrSerThrThrSerValGlnProSer 215  
Qy 1046 TAGTCTGCTAAGCGTCCGAGGAAAGACATTCGCGAGATCGGATCCGTAAGCAAGCCAAAG 1105  
Db 216 SerSerSerSerValProThr-----ThrSerAlaThrSerValAtgSerSerSer 232  
Qy 1106 GTGCTGCTATGGCGATGATTGCTCTCTCGAAAGAGGCTTAAAGCAAGGACCAAGG 1165  
Db 233 SerSerSerThrProIleProSerThrThrSerValGlnProSerSerSerSerAla 252  
Qy 1166 CCAACCGTGAGATGACTATTATTGACGAGACGCTCTCGGTGGTACGAGGTGCTGCTT 1225  
Db 253 ProThr-----ThrSerAlaThrSerValGlnProSerSerSerSerSer 266  
Qy 1226 TTGCACAGTTTGGCAAGCATGTCGCTACAAATCTGATGCTATCGCGGTATCTCAAAAT 1285  
Db 267 SerThrProIleProSer-----ThrThrSerValGlnProSerSerSerSer 282  
Qy 1286 CGCTTGACGACAACTACAAAGACTACCTCCCTCGCTGCGACCGTAAGGTGCTGCCGCCA 1345  
Db 283 SerAlaProThr-----ThrSerAlaThrSerValGlnProSerSerSerProPro 301  
Qy 1346 TGCTCGATATTGTACGCGCGGTATCCTCGCGCAAGCTCCCGGATATATTCAAGAATG 1405  
Db 302 IleSer-----SerThrIleSerValGlnProSerSer----- 312

Qy 1406 TAATCGACAAGAAATTCAAAGCGCACGACGAAGAAGTATGCAGACTTCGTATTTCGAAGA 1465  
Db 313 ---SerSerSerSerProThrThrThrSerThrThrValGlnProSerSerSerGlySer 331  
Qy 1466 GTGTGGTTCCTTATAGGACAAAGTTCCATGCCATGCTCAAGTCCATGCAACGAAGAAAGT 1525  
Db 332 AlaProThrThrSerAlaThrSerValGlnProSerSerSer-----Ser 346  
Qy 1526 TTGCCAAGGCTATCGAANAAGATCCGCGCAGTAGAGCTTTCCAAAGCGTAATAGCTGCTG 1585  
Db 347 SerProProIleSerSerThrIle----- 354  
Qy 1586 CTCGCGCTATTTCAGCGCGATGCGATGCGCAATGCTATGCAATTGAGAAGGCAAGCGTC 1645  
Db 355 -----SerValGlnProSerSerSer 361  
Qy 1646 TTTTCTTTTGGCGTTTTCGCTGAGATGTACCCCGGACGTCTCTCCGAGGATGCCAACT 1705  
Db 362 SerSerSerPro-----ThrThrSerThrThrSerValGlnProSerSerSerGlySer 379  
Qy 1706 TCACCATCGCTATGAGCTACGGCT-----CCATCAAGGGATATGAACCCGACGACGGTG 1759  
Db 380 AlaProThrThrSerAlaThrSerValGlnProSerSerSerSerSerValProThrThr 399  
Qy 1760 CCTGGTACAACATATCATACGACGAGGCGGTATTGGAGAACGAGGATCCTAAGAGCG 1819  
Db 400 SerAlaThrSerVal-----ArgSerSerSerSerSerSer 411  
Qy 1820 ATGAGTTTGGCTACAGG-----AGATATCTCGACTCTTCCGACACCAAAACT 1870  
Db 412 ThrProIleProThrThrThrSerValGlnProSerSerSerSerSerValProThrThr 431  
Qy 1871 ATGTCTCGTATCGCGAGAACGGTCAGCTCCATATCGCTTCTCTATCGAACAACGACATCA 1930  
Db 432 SerAlaThrSerValGlnThrSerSerSerSerSerSerProIleProSerThrThrSer 451  
Qy 1931 -----CGGCGGTAACTCCGCTAGCCCGCTAT 1957  
Db 452 ValGlnProSerSerSerSerSerAlaProThrThrThrSerValGlnProSer 471  
Qy 1958 TCGATAGAAGCGCGCTGATCGTCTGCTTTCGATGGCAACTGGGAAGCTATGAGTG 2017  
Db 472 SerSerSerSerProProIleSerSerThrIleSerValGlnProSer----- 487  
Qy 2018 GTGACATCGAGTTCCAAACCCGCTCGACGCGACAAATCAGGTGGACATCGCTACGTTTC 2077  
Db 488 ---SerSerSerSerSerProThrThrThrThrThrSerValGlnProSerSerSerGly 506  
Qy 2078 TCTTCATGATTGACAAATGGGGTCAGTCCCGCTCTCATCCAAAGAC 2125  
Db 507 SerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSer 522  
RESULT 5  
S48478  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)  
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIRO1  
C:Species: Saccharomyces cerevisiae  
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 12-Nov-1999  
C:Accession: S48478; A26877; S27281; JG6123  
R:Rowley, K.  
submitted to the EMBL Data Library, October 1994  
A:Reference number: S48478  
A:Accession: S48478  
A:Molecule type: DNA  
A:Residues: 1-1367 <ROW>  
A:Cross-references: GB:247047; EMBL:238061; NID:9603997; PID:9763364; GSPDB:GN000009;  
R:Yamashita, I.; Nakamura, M.; Fukui, S.  
J. Bacteriol. 169, 2142-2149, 1987  
A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.  
A:Reference number: A91831; MUID:87194600; PMID:3106330  
A:Accession: A26877  
A:Molecule type: DNA  
A:Residues: 1-242 <YAM>





```
Db 833 -----GluSer 834
QY 1604 ATCGATGGCAATGCTATGCTATGAGAGGCAAGCGTCTTTCTTTCGCGGTTTCG 1663
Db 835 SerSerValProValProThrProSerSerSerThrThrGluSerSerAlaProVal 854
QY 1664 GTGAGATGTACCCCGGAGCTCTGCGGAGCGATGCCAACT-----TCACCATGC 1714
Db 855 SerSerSerThrThrGluSerSerValAlaProValProThrProSerSerSerAsn 874
QY 1715 GTATGACTACGGCTCCTCAAGGGATATACCGGAGGAGCGTGCCT-----1762
Db 875 IleThrSerSerAlaProSer-----SerIleProPheSerSerThr 888
QY 1763 -----GGTACAACTATCATACGACAGCAGGCGGTATGGAGAACG 1804
Db 889 ThrGluSerPheSerThrGlyThrValThrProSerSerSerLysThrProGlySer 908
QY 1805 AGGATCCTAAGAGCGATGAGT-----TTGCGGTACAGGAGGAATA 1843
Db 909 GlnThrGluThrSerValSerSerThrThrGluThrThrIleValProThrLysThr 928
QY 1844 TCCTCGACCTCTCCCGCACCAAAACTATGTC-----1876
Db 929 ThrSerValThrProSerThrThrThrIleThrThrThrValCysSerThrGlyThr 948
QY 1877 -----GCTATGCGGAGAACCGCTCAGCTCCCATCGCTT 1909
Db 949 AsnSerAlaGlyGluThrThrSerGlyCysSerProLysThrValThr-----965
QY 1910 TCCTATGCAACAGACATACGCGGCGGTAACTCCGGTAGCCCGGTATTCGATAGAAGC 1969
Db 966 ThrValProThrThrThrThrSerValThrThrSerThrThrThrIleThrThr 985
QY 1970 CCGGTCTGATCGGCTTCTTCGATGCACTGGAGAGCTATGAGTGGTGCATCGAGT 2029
Db 986 ThrVal-----CysSerThrGlyThrAsnSerAlaGlyGluThrThrSerGly 1001
QY 2030 TCGAACCG-----ATCTGCGGCGCACAACTACGCGTGGACATCC 2068
Db 1002 CysSerProLysThrIleThrThrThrValProCysSerThrSer-----ProSer 1018
QY 2069 GCTACGTTCTCTTCATGATGACAAATGGGTGATGCGCCCGCTCATCC 2119
Db 1019 GluThrAlaSerGluSerThrThrThrSerProThrThrProValThrThr 1035

RESULT 6
S44784
C30C11.4 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 29-Sep-1999
C:Accession: S44784
R:Favell, A.D.
submitted to the EMBL Data Library, January 1993
A:Description: Sequence of the C. elegans cosmid C30C11.
A:Reference number: S44784
A:Accession: S44784
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-776 <Fav>
A:Cross-references: EMBL:L09634; NID:gl56220; PIDN:AAA27967.1; PID:gl56223
C:Genetics:
C:Introns: 36/2; 145/3; 255/1; 417/2; 732/3
C:Superfamily: heat shock protein 91

Alignment Scores:
Pred. No.: 0.117 Length: 776
Score: 143.50 Matches: 118
Percent Similarity: 37.48% Conservative: 87
Best Local Similarity: 21.57% Mismatches: 189
Query Match: 3.76% Indels: 153
DB: 27 Gaps: 2
```

```
US-10-008-355-1 (1-2139) x S44784 (1-776)
QY 113 ATCTGGATCGAATGCGTAGCTCGGTTTACGCTCCGTTGGATTGCTCTACAGTTTCG 172
Db 236 IleTrpPheAsp-----AlaLeuIleArgGluHisPheArgLysGluPheLys 251
QY 173 ACAAGCCGTCATG-----CCAATGCCGTGGTTATCTTCGGTGGCGGAT 217
Db 252 ThrLysTyroGlyIleAspAlaIaThrSerProArgProTrpLeuArgLeuLeuAspGlu 271
QY 218 GTACCGGTATACAGTCTCCGATCAGGCGCTGATCTTTACCAACCCACTCGGGATACG 277
Db 272 CysGluArgValLysLysGlnMetSerAlaAsnGlnThrProIle-ProLeuAsnIleG1 291
QY 278 GTGCTATCCAGACGCAAGCAGCGTGGATCAGCACTATCTGCGGAGTGGTTTCTTC 337
Db 291 uCysPheMetGluAspLysAspValThrGlyLysMetGln-ArgGlnGluPhe-----308
QY 338 GCAGGATGGGTGAGGAGCTTCCGATTCCG-----366
Db 309 -----GluAspLeuAlaAlaProIlePheAsnArgIleLysGlnValLeuIleA 325
QY 367 -----GCTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAAACGACA 412
Db 325 snLeuPheAlaAspGlyValSerIleLys-----ProGluGluIleAspG 340
QY 413 AGTGAAGAGACAGCTCAAGGTTATCACTGACGAGATGGAGCGTCTCGGCAACACTCAGG 472
Db 340 IuIleGlu-----IleValGlyGlySerSerArgIleProMetIleArg 355
QY 473 AGGTATGCCAAGAACTGCCCAAAAGAA-----AATGCAGACGAG-----513
Db 355 IuIleValLysAspLeuPheGlyLysGluProLysThrThrMetAsnGlnAspGluAlav 375
QY 514 -----AACCAACTCTGATCGTAGACCTTCTTATTTCCAAACCAAGAT 556
Db 375 alAlaArgGlyAlaAlaMetGlnCysAlaIleLeuSerProThrPheArgValArgGluP 395
QY 557 ACTTCTCATCTGCTACGATGTATTCAGGAC-----GTTCTGTATGTAT 601
Db 395 he-----AlaIleLysAspThrGlnProTyroArgIleArgLeuSerT 409
QY 602 TTGCTCTCCAGCTCTCTAGTAGTTCGAGGCGATACGACAACTGATGTGGCGGC 661
Db 409 rp-----AsnSerThrGlyGluAsnGlyGlyGluAsnAspValPheSer---ProA 425
QY 662 GTCACACGCGGACTTTCAGC-----GTATTCCGCGGTATGCCGCGCCGACACCGGC 715
Db 425 rAspGluValProPheSerLysLeuValSerLeuLeuArgSerGlyPropheAsnValG 445
QY 716 CGGCGGAATACAGCAAGACAAAT---AAACCTATAAGCCCGTTTACTTCCGTCGCGTAT 772
Db 445 luAlaHistyAlaGlnProAsnValValProHisAsnGlnValHisIleGlySerTrpL 465
QY 773 CCATCAAGGCTACAAAGCTGACGACTATGCCATCAGCTTTCGCGGCGACATCGG 832
Db 465 ysValAsnGlyAlaArg-----ProGlyAlaAsp- 474
QY 833 ATCGTACTCTACTTCTTGGGTGTGGAAGATCGTATCGAAACGAGAACAACTCTCTCA 892
Db 475 -----GlyGlyAsnGlnLysVal-----LysV 482
QY 893 TCGAAGTTGCGGGTATCAAGCAAGCATCTGGAAGAGACCCATGACCGCAGATCAGGCTA 952
Db 482 alLysValArgValAsnProAspGlyIlePheThrIleAla-----495
QY 953 CCCGTATCAATATCCAGCAGAGTATGCTCAGAGTGCCTAACTATTGGAAGAAATTCGATCG 1012
Db 496 -----SerAlaThrMetTyroGluProArgIleV 505
QY 1013 GTATGAACCGCGTCTCGCTCTTGAGTGTATAGGTGCTAAGCGTCCGAGGAAAGAG 1072
Db -----
```





QY 754 GTTACTTCGTCGCGGTATCCATGCAAGGCTACAAAGCTGACGACTATGCCATGACCATC 813  
 Db 181 rThrAspSerGlnIleSerLysGluGluGlyThrLeuAlaGlnThr 198  
 QY 814 GGTTCCTCCGGCAGTACGATCGCTACCTCCTCTCTTGGGGTGGGAAGATCGTATCGAA 873  
 Db 198 rIleSerGluThrThrArgIleAlaGlnMetValThrArgValSerGlnIleSerIle 218  
 QY 874 AAGCAAGCAATCCTCTATCGAAGTTTCGGGTATCAAGCAAGGCATC-----TGGAAAGGA 929  
 Db 218 eThrAlaAlaSerThrIleAspGlyPheSerSerGluSerThrClnThrAspPheSerAs 238  
 QY 930 AGCATGAGCGGAGATCAGCTACCCGTATCAATAATATGCCAGCAAGTATGCTCAGAGTGC 989  
 Db 238 nThrValSerPheGluAsnSerValGluGluGlyAlaMetSerLysSerGlnLeuSe 258  
 QY 990 TAACTATTGGAAGATTCGATCGGTATGACCGCGGTCTCGCTCTTTCACGTGATAGG 1049  
 Db 258 rGluSerTySerSerSerThrValTySerGly-----G 271  
 QY 1050 TCGTAAGCGTCCGAGGAAGACATTCGCAGACTGTCGTAAGAACGGCAAGAGTGC 1109  
 Db 271 yGluSerThrAlaAsp-----LysThrSerSerSerPr 282  
 QY 1110 TGCTATGGCGATGATATGTCTCTCTCGAAAGGCTTATAAGGAAGGACCAAGGCCAA 1169  
 Db 282 oIleThrSerPheSerSerSerTySerGlnThrThrSerThrGluThrSerGluSerSe 302  
 QY 1170 CCGTGAGATGACITATTGTAGC-----GAGACGCTCTTGGGTGACCGAGGTGGTTCG 1223  
 Db 302 rArgValAlaValGlyValSerArgProSerSerIleGlnThrThrSerIleAspSe 322  
 QY 1224 TTTTGCAAGTTTCCCAAGCATGGCTGCAAAATCCTGATGCTCATGCCGT----- 1275  
 Db 322 rPheSerMetSerGluValGluLeuSerThrTyThrAspLeuSerAlaGlyAsnTyPr 342  
 QY 1276 -----ATCCTCAAAATCGCTTGACGACAAGTACAAAGACTACCTCCCTC 1319  
 Db 342 oAspGlnGluLeuIleVal-AspArgProAlaThrSerSerThrAlaGluThrSerSerG 362  
 QY 1320 GCTGACCGTAAGGTGTCGCCGCGATGCTGATATTGTACCGCGGCGTATCCTCGCGGA 1379  
 Db 362 luAlaSer-----GlnGlyValSerArgGluS 371  
 QY 1380 CAAGTCCCGCATATATTCAAGATGTAATGACGACAAAGAAAT-----TCAAAGGCGA 1430  
 Db 371 eAsnThrPheAlaValSerSerIle-----SerThrThrAsnPheIleValSerSerAla 390  
 QY 1431 CACGAAGAAGTATCAGACTTCGTATTTCGACAAAGAGTG-----T 1469  
 Db 390 eAspThrValValSerThrSerSerThrAsnThrValProTySerSerValHisSerT 410  
 QY 1470 GGTTCCTTATAGCAGACAGTTCATGCTCATGCTCAAGTCCATGACGACGAAAGATTG- 1528  
 Db 410 hrPheValHisAlaThrSer-----SerSerThrTyIleSerSerSerLeut 426  
 QY 1529 -----CCAAGCTCTCGAAGAAGATCCGGCAGTAGAGTCTTCCCAAGACGCGTAATAGC 1580  
 Db 426 yrSerSerProSerLeuSerAlaSerValSerSer---HisPheGlyValAla----- 442  
 QY 1581 TGCTGCTCGCGCTATTACGCGCGATGCGATGCGCAATGCCATTATGCCATTGAAAGGCGAA 1640  
 Db 443 -----ProPheProSerAlaTyIleS 450  
 QY 1641 CCGTCTTTTCTTTCGCGTTGCGTGAGATGTACCCCGGAGCTGCTCTGCGGACGATGC 1700  
 Db 450 ePheSerSerValProValAlaValSerSerThr-----TyrT 463  
 QY 1701 CAACTTCACCATGCGTATGACGTACGCTCCATCAAGGGATATGAACGCGAGCGGTGC 1760  
 Db 463 hrSerSerProSerAlaSerValValProSerAlaTyAlaSerSerProSerValP 483  
 QY 1761 CTGTACA-----ACTATCATACGACAGCGGCGGTATTGGGAAGCA 1805

Db 483 roValAlaValSerSerThrTyThrSerSerProSerAlaProAlaAlaIleSerSerT 503  
 QY 1806 GGATCTTAAGACGATGATGTTGCGGTACAGGAGAATATCCTCGACCTCTTCCGACCAA 1865  
 Db 503 hrTyThrSerSerProSerAlaProValAlaValSerSerThrTyThrSerSerProS 523  
 QY 1866 AAACATATGGTGGTATGCCGAGAACGTCAGCTCC-----ATAT 1904  
 Db 523 eAlaProAlaAlaIleSerSerThrTyThrSerSerProSerAlaProValAlaValS 543  
 QY 1905 CGCTTCTCTAT-----CGAACACGACATCAC 1931  
 Db 543 eSerThrTyThrSerSerProSerAlaProAlaAlaIleSerSerThrTyThrSerS 563  
 QY 1932 GGGCGGTAACTCCGGTAGCCCGGTATTTCGATAGAACGGCGTCTGATCGGTCTTGGTTT 1991  
 Db 563 eProSerAlaProValAlaValSerSerThrTyThrSerSerProSerAlaProVala 583  
 QY 1992 CGATGGCAACTGGGAAGCTATGAGTGGTACATCGAGTTCGAACCCGATCTGCAGCGCAC 2051  
 Db 583 lAlleSerSer-----ThrTyThrSerSerProSerValProValAlaValS 599  
 QY 2052 AATCAGCGTGGACATCC-----GCTACGTTCTTTCATGATGACAAATG 2096  
 Db 599 eSerThrTyThrSerSerProSerAlaProAlaAlaIleSerSerThrTyThrSerS 619  
 QY 2097 GGGTCAGTCCCGCTCTCATCCAGAGC 2125  
 Db 619 eProSerAlaProValAlaValSerSer 628  
 RESULT 9  
 A53102  
 alpha-2-macroglobulin receptor precursor - chicken  
 N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein rece  
 C:Species: Gallus gallus (chicken)  
 C>Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 22-Jun-1999  
 C:Accession: A53102  
 R:Nimf, J.; Stifani, S.; Billous, P.T.; Schneider, W.J.  
 J. Biol. Chem. 269, 212-219, 1994  
 A:Title: The somatic cell-specific low density lipoprotein receptor-related protein o  
 A:Reference number: A53102; MUID:94103212; PMID:7506255  
 A:Accession: A53102  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-4543 <NIM>  
 A:Cross-references: GB:X74904; NID:g438006; PIDN:CAA52870.1; PID:g438007  
 C:Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-assoc  
 d protein.  
 C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-bind  
 C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glyco  
 F:1-17/Domain: signal sequence #status predicted <SIG>  
 F:18-342/Domain: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>  
 F:18-342/3943-4543/Product: alpha-2-macroglobulin receptor #status predicted <MAT>  
 F:29-66/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
 F:74-110/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F:117-150/Domain: EGF homology <EG1>  
 F:156-190/Domain: EGF homology <EG2>  
 F:200-241/Domain: LDL receptor YWTD-containing repeat homology <YW01>  
 F:242-283/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
 F:294-336/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
 F:337-380/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
 F:381-422/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
 F:423-470/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
 F:480-521/Domain: EGF homology <EG3>  
 F:573-615/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
 F:616-661/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
 F:662-712/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
 F:713-754/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
 F:755-797/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
 F:805-840/Domain: EGF homology <EG4>  
 F:852-888/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
 F:893-929/Domain: LDL receptor ligand-binding repeat homology <LDL4>

F:934-969/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
F:974-1009/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
F:1013-1049/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
F:1060-1095/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F:1102-1138/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
F:1143-1180/Domain: LDL receptor ligand-binding repeat homology <LDL1A>  
F:1183-1219/Domain: EGF homology <EG5>  
F:1225-1259/Domain: EGF homology <EG6>  
F:1267-1306/Domain: LDL receptor WTD-containing repeat homology <YW12>  
F:1307-1353/Domain: LDL receptor WTD-containing repeat homology <YW13>  
F:1354-1396/Domain: LDL receptor WTD-containing repeat homology <YW14>  
F:1397-1443/Domain: LDL receptor WTD-containing repeat homology <YW15>  
F:1444-1486/Domain: LDL receptor WTD-containing repeat homology <YW16>  
F:1487-1529/Domain: LDL receptor WTD-containing repeat homology <YW17>  
F:1538-1576/Domain: EGF homology <EG7>  
F:1581-1624/Domain: LDL receptor WTD-containing repeat homology <YW18>  
F:1625-1667/Domain: LDL receptor WTD-containing repeat homology <YW19>  
F:1668-1711/Domain: LDL receptor WTD-containing repeat homology <YW20>  
F:1712-1751/Domain: LDL receptor WTD-containing repeat homology <YW21>  
F:1752-1794/Domain: LDL receptor WTD-containing repeat homology <YW22>  
F:1795-1842/Domain: LDL receptor WTD-containing repeat homology <YW23>  
F:1846-1882/Domain: EGF homology <EG8>  
F:1930-1972/Domain: LDL receptor WTD-containing repeat homology <YW24>  
F:1973-2015/Domain: LDL receptor WTD-containing repeat homology <YW25>  
F:2016-2059/Domain: LDL receptor WTD-containing repeat homology <YW26>  
F:2060-2101/Domain: LDL receptor WTD-containing repeat homology <YW27>  
F:2102-2147/Domain: LDL receptor WTD-containing repeat homology <YW28>  
F:2155-2190/Domain: EGF homology <EG9>  
F:2195-2237/Domain: LDL receptor WTD-containing repeat homology <YW29>  
F:2247-2288/Domain: LDL receptor WTD-containing repeat homology <YW30>  
F:2338-2382/Domain: LDL receptor WTD-containing repeat homology <YW31>  
F:2383-2423/Domain: LDL receptor WTD-containing repeat homology <YW32>  
F:2424-2467/Domain: LDL receptor WTD-containing repeat homology <YW33>  
F:2476-2511/Domain: EGF homology <EG10>  
F:2518-2555/Domain: LDL receptor ligand-binding repeat homology <LDLB>  
F:2560-2594/Domain: LDL receptor ligand-binding repeat homology <LDLC>  
F:2599-2633/Domain: LDL receptor ligand-binding repeat homology <LDLD>  
F:2646-2682/Domain: LDL receptor ligand-binding repeat homology <LDLE>  
F:2690-2724/Domain: LDL receptor ligand-binding repeat homology <LDLF>  
F:2732-2767/Domain: LDL receptor ligand-binding repeat homology <LDLG>  
F:2772-2810/Domain: LDL receptor ligand-binding repeat homology <LDLH>  
F:2816-2851/Domain: LDL receptor ligand-binding repeat homology <LDLI>  
F:2856-2895/Domain: LDL receptor ligand-binding repeat homology <LDLJ>  
F:2902-2936/Domain: LDL receptor ligand-binding repeat homology <LDLK>  
F:2941-2977/Domain: EGF homology <EG11>  
F:2983-3018/Domain: LDL receptor WTD-containing repeat homology <YW34>  
F:3026-3065/Domain: LDL receptor WTD-containing repeat homology <YW35>  
F:3066-3110/Domain: LDL receptor WTD-containing repeat homology <YW36>  
F:3111-3153/Domain: LDL receptor WTD-containing repeat homology <YW37>  
F:3154-3197/Domain: LDL receptor WTD-containing repeat homology <YW38>  
F:3198-3238/Domain: LDL receptor WTD-containing repeat homology <YW39>  
F:3239-3281/Domain: LDL receptor WTD-containing repeat homology <EG12>  
F:3291-3327/Domain: EGF homology <EG13>  
F:3331-3366/Domain: LDL receptor ligand-binding repeat homology <LDLL>  
F:3371-3405/Domain: LDL receptor ligand-binding repeat homology <LDLM>  
F:3410-3445/Domain: LDL receptor ligand-binding repeat homology <LDLN>  
F:3450-3486/Domain: LDL receptor ligand-binding repeat homology <LDLO>  
F:3491-3528/Domain: LDL receptor ligand-binding repeat homology <LDLP>  
F:3533-3567/Domain: LDL receptor ligand-binding repeat homology <LDLQ>  
F:3572-3606/Domain: LDL receptor ligand-binding repeat homology <LDLR>  
F:3610-3644/Domain: LDL receptor ligand-binding repeat homology <LDLS>  
F:3651-3687/Domain: LDL receptor ligand-binding repeat homology <LDLT>  
F:3692-3728/Domain: LDL receptor ligand-binding repeat homology <LDLU>  
F:3738-3774/Domain: LDL receptor ligand-binding repeat homology <LDLV>  
F:3783-3820/Domain: EGF homology <EG14>  
F:3826-3858/Domain: EGF homology <EG15>  
F:3866-3909/Domain: LDL receptor WTD-containing repeat homology <YW40>  
F:3910-3968/Domain: LDL receptor WTD-containing repeat homology <YW41>  
F:3943-4543/Domain: alpha-2-macroglobulin receptor 85K chain #status predicted <85K>  
F:3943-4420/Domain: 85K chain extracellular #status predicted <EXT>  
F:3969-4011/Domain: LDL receptor WTD-containing repeat homology <YW42>  
F:4012-4055/Domain: LDL receptor WTD-containing repeat homology <YW43>  
F:4056-4098/Domain: LDL receptor WTD-containing repeat homology <YW44>

F:4099-4141/Domain: LDL receptor WTD-containing repeat homology <YW45>  
F:4150-4181/Domain: EGF homology <EG16>  
F:4199-4230/Domain: EGF homology <EG17>  
F:4235-4266/Domain: EGF homology <EG18>  
F:4271-4302/Domain: EGF homology <EG19>  
F:4307-4338/Domain: EGF homology <EG20>  
F:4343-4373/Domain: EGF homology <EG21>  
F:4376-4408/Domain: EGF homology <EG22>  
F:4421-4443/Domain: transmembrane #status predicted <TMW>  
F:4444-4543/Domain: Intracellular #status predicted <INT>  
F:116,138,187,276,359,448,731,926,1048,1152,1153,1193,1216,1305,1509,1556,1573,1614,1  
3485,3659,3786,3837,3952,4074,4124,4178,4278/Binding site: carbohydrate (Asn) (coval  
F:165,2995/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F:2955/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted

## Alignment Scores:

Pred. No.:	0.511	Length:	4543
Score:	139.00	Matches:	139
Percent Similarity:	31.52%	Conservative:	69
Best Local Similarity:	21.06%	Mismatches:	196
Query Match:	3.64%	Indels:	256
DB:	1	Gaps:	37

US-10-008-355-1 (1-2139) x A53102 (1-4543)

Qy 268 TCGGATACGGTGTATCCAGACGCAAGCGGTGGATCAGACTATCTCGCGATGGT 327

Db 3765 CysGlyAspGlySerAspGluLysSerCysSerHisGluHisLysSerTyr---AspCys 3783

Qy 328 TTCGTTCTCCACGATG---GGTGAGGAGCTTCG---ATTCGGGGCTTTTCGCTGAAG 381

Db 3784 MetThrAsnThrThrMetCysGlyAspGluAlaGlnCysIleGlnAlaGlnSerSerThr 3803

Qy 382 TATCTG-----CGCAAGATCCTGAAGTACGGACAGGAGTAGAGGACGACTCAAG 432

Db 3804 TyrCysThrCysArgArgGlyPheGlnLysValProAspLysAsnSerCysGln----- 3821

Qy 433 GGTATCACTGACGAGATCGAGCTCTCGCC----- 462

Db 3822 -----AspValAsnGluCysLeuArgPheGlyThrCysSerGlnLeuCysAsnAsn 3838

Qy 463 ---AAAGCTCAGGAGGTATGC-----CAAGAACTGGCCCAAAAAGAAAT----- 504

Db 3839 ThrLysGlySerHisValCysSerCysAlaLysAsnPheMetLysThrAspAsnMetCys 3858

Qy 505 -----CGACGAGAGCAACTCTGCATC 528

Db 3859 LysAlaGluGlySerGluHisGlnIleLeuTyrIleAlaAspAspAsnLysIleArgSer 3878

Qy 529 GTAGAGCCTTCTATTCACCAACCAAGTAATCTTCCTCATCTGCTACGATGTATTCAAGGAC 588

Db 3879 MetTyrProPheAsnProAsnSerAlaTyr----- 3888

Qy 589 GTTCGTATGGTATTTCCTCCAGCTCTAGTAGTAAGTTCGGAGCGCATACGAC--- 645

Db 3889 -----GluProAla-----PheGlnGlyAspGluAlaVal 3898

Qy 646 -----AACTGGATG 654

Db 3899 ArgIleAspAlaMetAspIleTyrValLysGlyAsnLysIleTyrTrpThrAsnTrp--- 3917

Qy 655 TGGCGCGCTCAGCGGGCGACTTCAGCGTATTCCGCGTGTATGCGCGT----- 702

Db 3918 -----HisThrGlyArgIleSerTyrCysGluLeuProAlaSerSerAlaAlaSer 3934

Qy 703 ---GCCGACACCGG----- 714

Db 3935 ThrAlaSerAsnArgAsnArgGlnIleAspGlyGlyValThrHisLeuAsnIleSer 3954

Qy 714 ----- 714

Db 3955 GlyLeuLysMetProArgGlyIleAlaValAspTrpValAlaGlyAsnIleTyrTrpThr 3974



```
QY 715 -----CCGGCCGAATACAGCAAGGACATAAACCCTATAAG 750
Db 3975 AspSerGlyArgAspValIleGluValAlaGlnMetLysGlyGluAsnArg----- 3991
QY 751 CCGGTTTACTTCGCTGCCGTATCCATCGAAGGCTACAGGCTACGACTATGCCATGACC 810
Db 3992 -----LysThrLeuIleSerGlyMetIleAspGluProHisAlaIleVal 4006
QY 811 ATCGGTTTCCGGCCAGTACGGATCGCTACCTCTCTTGGGGTGTGGGAAGATCGTATC 870
Db 4007 ValAsp---ProLeuArgGlyThrMetTyrTrpSerAspTrpGly----- 4020
QY 871 GAAACGAGAACATCTCTATCGAAGTTCCGGGTATCAAGCAAGCATCTGGAAGGAA 930
Db 4021 -----AsnHisProLysIleGluThrAlaAlaMet---AspGlyThrLeuArgGlu 4036
QY 931 GCCATGAGCCACATCAGGCTACCGGTATCAAAATAT-----GCCAGCAAGTAT 978
Db 4037 ThrLeuValGlnAsp-----AsnIleGlnTrpProThrGlyLeuAlaValAspTyr 4053
QY 979 GCTCAGAGTCTAACTATTGGAAGATTCGATCGGTATGAACCGCGTCTCGCTCTT 1038
Db 4054 HisAsnGluArgLeuTyrTrpAlaAsp-----AlaLysLeu 4065
QY 1039 GACGTATAGTCGTAAAGCGTCGGAGGAAGAGCATTCGCAGCTGGATCCGTAGAAC 1098
Db 4066 SerValIleGlySer-----IleArgLeuAsn 4074
QY 1099 GGC----- 1101
Db 4075 GlyThrAspProValValAlaIleAspAsnLysLysGlyLeuSerHisProPheSerIle 4094
QY 1102 -----AAGAGTGTGTCTATGGC-----GATGATTATGTCTTCTCGAAAGGCT 1146
Db 4095 AspIlePheGluAspTyrIleTyrGlyValThrTyrIleAsnAsnArgIlePheLysIle 4114
QY 1147 TATAAGGAAGGCCAAGGCCAACCGGTGAGATGACTTATTTTGACGAGACGCTCTTCGGT 1206
Db 4115 HisLysPheGlyHisLysSer-----ValThrAsnLeuThrSerGlyLeuAsnHis 4131
QY 1207 GGTACCGAGCTGCTCGTTTGGCACAGTTTCCCAACGCATTTGGCTACAAATCCTGATGCT 1266
Db 4132 AlaThrAspValValLeuTyrHisGlnTyrLysGlnProGluValThrAsnPro----- 4149
QY 1267 CATGCCGATCTCCTCAAAATCGCTGACGACAAGTACAAAGACTACCTC----- 1314
Db 4150 -----CysAspArgLysLysCysGlnTrpLeuCysLeuLeuSer 4162
QY 1315 CCTCGCTGACCGTAAGGTGCTGCCCGCATCTCGATATTTGACCGCGC----- 1366
Db 4163 ProSer-GlyProValCysThrCysProAsnGlyLysArgLeuAspAsnGlyThrCysVa 4182
QY 1367 -----GTATCCTCGCCAGACGTCCTCGATATTTCAAGATGTAATCGACAAGATTC 1422
Db 4182 LeuIleProSerProThrAlaSerAlaVal-ValProThrThrAspThrCysAspLeuV 4202
QY 1423 AAAGGCGACACGAAGAAGTATGACAGATTCGATTCGACAAGAGTGTTGCTTATAGC 1482
Db 4202 alCysLeuAsnGlyGlySerCysPheLeuAsnAlaArg-----L 4215
QY 1483 GACAAGTTCCATGCCATGCTCAAGTCCATCGACAAGGAAAGTTTGCCAA----- 1532
Db 4215 ysGlnAlaLysCysArgCysGlnProArgTyrAsnGlyGluArgCysGlnIleAsnGlnC 4235
QY 1533 -----GCCTATCAGAAAGATCCGGC-----AGTAGAGCTTTCACAGAGCGTA 1575
Db 4235 ysSerAspTyrCysGlnAsnGlyCysLeuCysThrAlaSerProSerGlyMetProThrC 4255
QY 1576 ATAGCTCTGCTCGCGCTATTTCAGGC---CGATCGCATGCCAA---TGC-----CTAT 1623
Db 4255 ysArgCysProThrGlyPheThrGlySerArgCysAspGlnGlnValCysThrAsnTyrC 4275
QY 1624 GCCATTGAGAAGGCCAAGCGCTCTTT-----CTTTGCCGGTTTCGCT 1665
```

```
Db 4275 ysHisAsnAsnGlySerCysThrValAsnGlnGlyAsnGlnProAsnCysArg----- 4292
QY 1666 GAGATGTACCCCGAGCGTCTCTCCGAGCGATGCCAAGCTTCCACCATCGCTATGAGCTAC 1725
Db 4293 -----CysProProThrPheIleClyAspArgCysGlnTyrGlnGlnCysPheAsnTyrC 4311
QY 1726 GGCTCATCAAGGATATGAACCGCAGGACGGTGCCTGGTACAACTATCATACAC 1781
Db 4311 ysGluAsnAsnGlyVal-----CysGlnMetSerArgAsp 4322
RESULT 10
S53362
mucin 5AC (clone JER47) - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1996 #sequence_revision 07-feb-1997 #text_change 08-Oct-1999
C:Accession: S53362; S71065
R:Guyonnet-Duperat, V.; Audie, J.P.; Debailleul, V.; Laine, A.; Buisine, M.P.; Galleg
Biochem. J. 305, 211-219, 1995
A:Title: Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich d
A:Reference number: S53361; MUID:95126907; PMID:7826332
A:Accession: S53362
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-477 <GUY>
A:Cross-references: EMBL:Z34277
A:Experimental source: clone JER47
R:Porchet, N.
Submitted to the EMBL Data Library, June 1994
A:Reference number: S71065
A:Accession: S71065
A:Molecule type: mRNA
A:Residues: 1-211, 'S', 213-224, 'AR', 227-259, 'S', 261-477 <POR>
A:Cross-references: EMBL:Z34277; NID:9563374; PIDN:CAA84031.1; PID:9563375
A:Experimental source: clone JER47
C:Genetics:
A:Gene: GDB:MUC5AC
A:Cross-references: GDB:454136; OMIM:158373
A:Map position: l1p15.5-l1p15.5
A:Keywords: glycoprotein; tandem repeat

Alignment Scores:
Pred. No.: 0.226 Length: 477
Score: 138.00 Matches: 74
Percent Similarity: 41.18% Conservative: 31
Best Local Similarity: 29.02% Mismatches: 94
Query Match: 3.61% Indels: 56
DB: 2 Gaps: 11

US-10-008-355-1 (1-2139) x S53362 (1-477)
QY 1413 CAAGAAATTCAAAGGCCACACGAGAGATATGCAGACTTCGTATTCGACAGAGTGTGGT 1472
Db 57 GluLysIleCysArgArgProGluGluIleThrArgLeuGlnCysArgAlaGlu----- 74
QY 1473 TCCTTATAGCGACAAGTTCCTCATGCCATGCTCAAGTCTCATGACAA-----GGAAAAGTT 1526
Db 75 -----SerHisProGluValSerIleGluHisLeuGlyGlnVal 87
QY 1527 TGCCAAGGCTATCGAGAAGATCGCGCAGTAGAGCTTTTCCAAGAGCGTAATAGTGTGTC 1586
Db 88 ValGln-----Cys 90
QY 1587 TCCGGCTTTCAGGCCCATCGGATGGCCAATGCCATATGCCATTGAGAAAGGCCAGCGTCT 1646
Db 91 SerArgGluGluGly-----LeuValCysArgAsnGlnAspGlnGlnGly 105
QY 1647 TTTCTTTGCCGGTTTGGTGTGAGATGTACCCCGAGCGTCTCGCGAGCGATGCCA----- 1702
Db 106 --ProPheLysMetCysLeuAsnTyrGluValArgValLeuCysCysGluThrProLysG 125
QY 1703 -----ACTTCACCATGCGTATGAGTACGGTCTCCATCAAGGATATGAACCGCA 1751
Db 1703 -----Cys 90
```



Db 125 lYcysProValThrSerThrProVal-----ThrAlaProSerThrProSerGlyArgA 143  
Qy 1752 GGACGGTCCGTGGTACAACTATACACAGCAAGCGCGTATTGG-----AGAA 1802  
Db 143 laThrSerPro-----ThrGlnSerThrSerSerTrpGlnLysSerArgT 158  
Qy 1803 CGAGGATCCATAGACGGATGAGTTGGCGTACAGGAGAAATATCTCGACCTCTTCGCGAC 1862  
Db 158 hrThrThrLeuValThrThrSerThrThrProGlnThrSerThrThrSerAlap 178  
Qy 1863 CAAAACATATGGTCCGTATCGGAGACGGTCAGCTCATATCGCTTCCTTCATCGAACAA 1922  
Db 178 roThrThrSerThrThrSerAlaProThrThrSerThrSerAlaProThrThrSert 198  
Qy 1923 CGACATACGGCGGTAACCTCCGGTA-----GCCCG-----TATTTCGATAAGACGGCG 1973  
Db 198 hrThrSerThrProGlnThrSerThrSerAlaProThrThrSerThrThrSerAlap 218  
Qy 1974 TCTGATCGGTCTGCTTCGATGCACTGGGAAGCTA-----TGAGTGGTGACAT 2024  
Db 218 roThrSerThrThrSerArgAlaThrThrSerIleSerAlaProThrThrSert 238  
Qy 2025 CGAGTTCGAACCGATTCGACGGCACAATCAGGTGGACATCGCTACGTTCTCTTCAT 2084  
Db 238 hrThrSerPheProThrThrSerThrThrSerAlaThrThrSerThrThrSerAlap 258  
Qy 2085 GATTGACAAATGGGTACGTCCCGCTCATCTCAAGAGC 2125  
Db 258 roThrThrSerThrThrSerThrProGlnThrSerLysThr 271

RESULT 11  
T45025  
mucin MUC5B, tracheobronchial [imported] - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T45025  
R:Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.  
J. Biol. Chem. 272, 3168-3178, 1997  
A:Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat  
A:Reference number: 228899; MUID:97166151; PMID:9013550  
A:Accession: T45025  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3570 <DES>  
A:Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503  
A:Experimental source: placenta  
C:Genetics:  
A:Gene: MUC5B

Alignment Scores:  
Pred. No.: 0.631 Length: 3570  
Score: 137.00 Matches: 154  
Percent Similarity: 33.29% Conservative: 94  
Best Local Similarity: 20.67% Mismatches: 293  
Query Match: 3.59% Indels: 204  
DB: 2 Gaps: 33

US-10-008-355-1 (1-2139) x T45025 (1-3570)

Qy 65 CCAAGCCGACAAAGGCGATGGCTCCTCAACGAACTCAATCAGGAGAACTCGGATCGAA 124  
Db 2906 ProThrAlaThrSerSerLysAlaThrSerSerSerProArg-----ThrAla 2922  
Qy 125 TGGGTGAGCTCGGCTTTACGCTCCGTTGGATTCGCTACAGTTTCGACACGCGTCCA 184  
Db 2923 ThrThrLeuProValLeuThrSerThrAlaThrLysSerThrAlaThrSerValThrPro 2942  
Qy 185 TTGCCAATCGCGTGTATCTTCGTCGGCGGATGATCCG----- 223  
Db 2943 IleProSerThrLeuGlyThrThrGlyThrLeuProGluGlnThrThrProVal 2962  
Qy 224 GTATCAGAGTTCGATCGAGCGCTGATCTTTACCAACCACT---GCGGATACGGTG 280

Db 2963 AlaThrMetSerThrIleHisProSerSerThrProGluThrThrHisThrSerThrVal 2982  
Qy 281 CTATCCAGAGCAAAAGCAGCGGTGATCAGCACTATCTCGCGGATGGTTTCGTTCTC--- 337  
Db 2983 LeuThrThrLysAlaThrThrArgAlaThrSerSerThrSerThrProSerSerThrPro 3002  
Qy 338 ---GCAGCATGGGTGAGGAGCTCCGATTCGGGTCTTTCGCGTGAAGTATCTGCGCAAGA 394  
Db 3003 GlyThrThrTrpIle----- 3007  
Qy 395 TCGTGAAGGTAAACGACAGAGTACAGGACAGCTCAAGGGTATCAGTGACGAGATGGAGC 454  
Db 3008 -----LeuThrGluLeuThr 3012  
Qy 455 GTCTGCGCAAAAGCTCAGGAGGTATGCCAAGAACTGGCCAAAAGAAATGCAGACGAGA 514  
Db 3013 ThrAlaAlaThrThr-----AlaGlyThrGlyProThrAlaThrProSerSerThr 3030  
Qy 515 ACCAACTCTGCATCGTAGAGCTTTTCTATTCCAACAGCAATACTTCTCATCTCTACG 574  
Db 3031 ProGlyThrThrTrpIleLeuThrGluLeuThrThrAlaThrThrAlaSerThr 3050  
Qy 575 ATGTATTCAAGGAGCTTC-----GTATGGTATTTCCTCCTCCTCCAGCT 616  
Db 3051 GlySerThrAlaThrLeuSerSerThrProGlyThrThrTrpIleLeuThrGluPro--- 3069  
Qy 617 CTGTAGGTAACTTCGAGGCGCATACGGACA-----ACTGGATGTGGCGC 661  
Db 3070 -----SerThrThrAlaThrValThrAlaProProGlySerThrAlaThrAlaSer 3086  
Qy 662 GTCACAGGCGGACTTCACGCGTATTCCGCGGTATGCCGGTGCAGCAACAGCGCGCGC 721  
Db 3087 SerThrGlnAlaThrAlaGlyThrProHisValSerThrThrAlaThrThrProThrVal 3106  
Qy 722 AATCAGCAAGGACAATAAACCCCTATTAAGCCGTTTACTTCGCTGCCGTATCCATGCAAG 781  
Db 3107 ThrSerSerLysAlaThrProSerSerSerProGlyThrAlaThrAlaLeuProAlaLeu 3126  
Qy 782 GCTACAAGGCTGAGCACTATCCATGACATCGGTTCCTCCGGGCGAGTACGATCGCTACC 841  
Db 3127 ArgSerThrAlaThr-----ProThrAlaThrSerPheThrAlaIleProSer 3143  
Qy 842 TCACCTCTTGGG---GTGTGGAAGATCGTATCGAAACAGCAACAATCCTCGTATCGAAG 898  
Db 3144 SerSerLeuGlyThrThrTrpThrArgLeuSerGlnThrThrProThrAlaThrMet 3163  
Qy 899 TTCGCGGTATCAAGCAAGGCATCTGGGAAGGACCATCAGCGCAGATCAGCGTACCCTA 958  
Db 3164 SerThrAlaThrProSerSerThr-----ProGlu 3173  
Qy 959 TCAATATGCCAGCAAGTATGCTC-----AGAGTGTAACTATTGGAAGAATTCGATCG 1012  
Db 3174 ThrValHisThrSerThrValLeuThrThrThrAlaThrThrGlyAlaThrGlySer 3193  
Qy 1013 GTATGAACCGCGGTCTCGCTCGTCTTGACGTGATAGTGTGTAAGCGTCCG----- 1063  
Db 3194 ValAlaThrProSerSerThrProGlyThrAlaHisThrThrLysValProThrThrThr 3213  
Qy 1064 -----AGGAAGAGCATTCGACAGACTCGGATCCGTAAGAACGCGGAGAGTG 1108  
Db 3214 ThrThrGlyPheThrAlaThrProSerSerSerProGlyThrAlaLeuThrProVal 3233  
Qy 1109 CTGTCTATGGGATGTTGCTCTCTCGAAAGGCTTATAAGGAAGGAGGAGGAGGAGGCA 1168  
Db 3234 -----TrpIleSerThrThrThrThrProThrProThrThrThrPro 3246  
Qy 1169 ACCGTGAGATGACTTATTGACGAGACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTGT 1228  
Db 3247 Thr-----ThrSerGlySerThrValThrPro----- 3255  
Qy 1229 CACAGTTTGGCA-----ACGCAATTGGCTACAAATCTCTG----- 1261  
Db 3256 SerSerIleProGlyThrThrHisThrAlaArgValLeuThrThrThrThrThrVal 3275

QY 1262 -----ATGCTCATGCGGCTATCCTCAAAATCGC----- 1288  
 Db 3276 AlaThrGlySerMetAlaThrProSerSerSerThrGlnThrSerGlyThrProProSer 3295  
 QY 1289 -----TTGAGCACAAGTACAAAGACTACCTCCCTCCCTCGACC 1327  
 Db 3296 LeuThrThrAlaThrThrIleThrAlaThrGlySerThrThrAsnProSerSerThr 3315  
 QY 1328 GTAAGGTGTCGCCCGCATGCTGATATTGACGCGGCTATCCCTGCGG-----ACA 1381  
 Db 3316 ProGlyThrThrProIleProProValLeuThrSerMetAlaThrThrProAlaAlaThr 3335  
 QY 1382 AGCTCCCGCATATTCAAGAATGTAATGACAAAGAAATTCAAAGCGCAGCAAGAAAGT 1441  
 Db 3336 SerSerLysAlaThrSer-----SerSerProArgThrAlaThrThrLeuPro 3352  
 QY 1442 ATCCAGACTTCGTATCGCACAAGAGTGTGTTCTTATACGACAAAGT-----TCCATG 1495  
 Db 3353 ValLeuThrSerThrAlaThrLys-----SerThrAlaThrSerPheThrProIle 3369  
 QY 1496 CCATGCTCAAGTCCATGACGACGAAAGTTTGCCAAAGCTATCGAGAAAGATCCGCGAG 1555  
 Db 3370 ProSerSerThrLeuThrThr----- 3376  
 QY 1556 TAGAGCTTCCAAAGCGTAATAGCTGCTGCTCGCGCTATTCAGGCGGATGCGATGCGCA 1615  
 Db 3377 ----- 3379  
 QY 1616 ATCCCTATGCCATTGAGAGGGCAAGCGTCTTTCCTTTCGCGGTTTCGCTGAGATGTACC 1675  
 Db 3380 ValProAlaGlnThrThrThrProMetSerThrMetSerThrIleHisThrSerSerThr 3399  
 QY 1676 CCGGACGTGCTGCGCGAGCGATGCCAACTTACCATGCGTATGAGTACGCGTCCA--- 1732  
 Db 3400 ProGlu-----ThrThrHisThrSerThrValLeuThrThrThrAlaThrMet 3415  
 QY 1733 TCAAGGGATATGAACCGCAGGCGGTGCTGGTACAACTATATACATACACAGGCAAGGCG 1792  
 Db 3416 ThrArgAlaThrAsnSerThrAlaThrProSerSerThrLeu----- 3429  
 QY 1793 TATGGAGACGAGGATCCTTAAGAGCGATGAGTTGCCGTACAGGAGATATCCTCGACC 1852  
 Db 3430 ---GlyThrThrArgIleLeuThrGluLeuThrThrThrAlaThrThrThrAlaAlaThr 3448  
 QY 1853 -----TCTTCCGACCAAAAAC-----ATGGTCGCTATGCGCAGA 1888  
 Db 3449 GlySerThrAlaThrLeuSerSerThrProGlyThrThrThrIleLeuThrGluProSer 3468  
 QY 1889 ACGTCAGCTCCATATGCGTT-----TCCTATGGAACAACAGCATACCGGCGGTA 1939  
 Db 3469 ThrIleAlaThrValMetValProThrGlySerThrAlaThrThrSerSerThrLeuGly 3488  
 QY 1940 ACTCCGGTACCCCGTATTCGATGATGAAGACGCGCTGATCGCTCTGCTTTTCGATGGCA 1999  
 Db 3489 ThrAlaHisThrPro-----LysValValThrAlaMetAla 3500  
 QY 2000 ACTGGGAAGCTATGAGTGTGACATGAGTTGGAACCGCATCTGACGCGCACAATACAGC 2059  
 Db 3501 Thr-----MetProThrAlaThrAlaSerThrValProSerSerThrVal 3516  
 QY 2060 TGGACATCCGCTACG 2074  
 Db 3517 GlyThrThrArgThr 3521  
 RESULT 12  
 A60999  
 alpha-amylase (EC 3.2.1.1) precursor - Micrococcus sp. (strain 207)  
 C:Species: Micrococcus sp.  
 C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 15-Oct-1999  
 C:Accession: A60999  
 R:Kimura, T.; Horikoshi, K.  
 FEMS Microbiol. Lett. 71, 35-42, 1990

A:Title: The nucleotide sequence of an alpha-amylase gene from an alkalopsychrotroph  
 A:Reference number: A60999  
 A:Accession: A60999  
 A:Molecule type: DNA  
 A:Residues: 1-1104 <KIM>  
 A:Cross-references: GB:X55799; NID:g296762; PIDN:CAA39321.1; PTD:g296763  
 C:Function:  
 A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds  
 A:Pathway: glycogen/starch degradation  
 C:Keywords: glycogen; hydrolase; polysaccharide degradation  
 F:1-32/Domain: signal sequence #status predicted <SIG>  
 F:33-1104/Product: alpha-amylase #status predicted <MAT>

## Alignment Scores:

Pred. No.: 0.659 Length: 1104  
 Score: 133.50 Matches: 138  
 Percent Similarity: 34.94% Conservatives: 94  
 Best Local Similarity: 20.78% Mismatches: 259  
 Query Match: 3.49% Indels: 173  
 DB: 2 Gaps: 35

US-10-008-355-1 (1-2139) x A60999 (1-1104)

QY 421 GGACAGCTCAAGGTATCACTAGCAGATGAGCGTTCGCCAAGCTCAGGAGGTATGC 480  
 Db 384 GlyAspPheLysGlyIleThrAspLysLeuAspTyrLeu----- 396  
 QY 481 CAAGAACTGCCCAAAAGAAATGACAGCAGAGAACCACTCTGCATCGTAGACCTTTC 540  
 Db 397 -----AspGluLeuGlyValAsnThrIleTrpIle---SerProVal 409  
 QY 541 TATTCCAACAAACGAATACTTCCTCATCTACGATGATTTCAAGGACGTTCGTATGGTA 600  
 Db 410 ValGluAsnIleLysTyr-----AspValArgTyrTyr 420  
 QY 601 TTTGCTCTCCAGCTCTGTAGTAAGTTCGGGCGCATACGGACAACCTGG-----ATG 654  
 Db 421 GluThrSerGluProTyrTyrGlyTyrHisGlyTyrTrpAlaAsnAsnPheGlyGluLeu 440  
 QY 655 TGCGCGGTCACAGCGGCGACTTCAGCGTATTCGCG---GTGTATCGCGTCCGACAAAC 711  
 Db 441 AsnPro---HisPheGlyThrMetGluGluPheHisAspLeuIleAspGlyAlaHisAsp 459  
 QY 712 CGCGCGCGCCGAATACAGCAAGGACAATAAACCCCTATAAGCCGCTTACTTCGCTCGCGTA 771  
 Db 460 ArgAsnMetLysIleMetValAsp-----ValValVal 470  
 QY 772 TCATGCAAGGCTAC-----AAGGCTGACGACTATGCCATGACC-----ATCGGT 816  
 Db 471 AsnHisThrGlyTyrGlyLeuLysGluIleAspGlySerValThrAsnProProAlaGly 490  
 QY 817 TTCGCGGCGAGTACGGATCGCTACCTCCTCT----- 849  
 Db 491 TyrProSerAspAlaAspArgAlaArgPheSerAspLeuLeuArgGlnGlyAlaAspVal 510  
 QY 850 -----TGGGGTGTGGAAGATCGTATCGAAACAGCAGAAC 882  
 Db 511 GlyThrAspGluValValGlyGluLeuAlaGlyLeuProAspPheIleThrGluAspPro 530  
 QY 883 AATCCTCGTATCGAAGTTCGGGTATCAAGCAGGCGATCTGGAGGAGCAGCAGTACGCGCA 942  
 Db 531 AsnValArgLysGlnIleIleAsp---TrpGlnThrAspTrpIleGluLysAlaThrThr 549  
 QY 943 GATCAGGCTACCCGCTATCAAAATAT-----GCCAGCAAGTATGCTCAGAGTGT 990  
 Db 550 GluAsnGlyAsnThrIleAspTyrPheArgValAspThrValLysHisValGluAspAla 569  
 QY 991 AACTATTGGAAGATTCGATCGGTATGAACCGCGCTCTCGCTCTGCTTACGCTGATAGGT 1050  
 Db 570 ThrTrpMetGlnPheLysAsnAlaLeuThrGluLysMetProGluPheLysMetIleGly 589  
 QY 1051 CGTAAGCGTCCCGAGGAAAGAGCATTCGACAGCTGGATCCGTAAGACGCGCAAGAGTCT 1110  
 Db 1111

Db 590 GluAla-----TrpGlyAlaLysValaspAsnThrLeu 600  
QY 1111 GTCTAT-----GCCGATGATTCTCTCTCCAAAAGGCTTATAAGGAGGAGCC 1161  
Db 601 GlyTyrLeuGluThrGlyThrMetaspSerLeuLeuLeuLeuGlyPheLysGluThrAla 620  
QY 1162 AAG-----GCCAACCGTGAGATGACGATTTATTTGAGCGAG 1194  
Db 621 ArgSerPheValAsnGlySerLeuGluAlaAlaAsnAlaSerLeuThrAlaArgAsnAla 640  
QY 1195 ACGCTCTCGGTGACCGAGGCGGTTCGTTT-----GCACAGTTT 1236  
Db 641 LysLeuAspAsnThrAlaThrLeuGlyGlnPheLeuGlySerHisAspGluGluGlyPhe 660  
QY 1237 GCCAACGATTGGCTACAAAT----- 1257  
Db 661 LeuHisSerLeuAlaGlyAspLysGlyLysLeuGlnValAlaAlaThrLeuGlnAlaThr 680  
QY 1258 -----CCTGATGCTCATGCCGATCCTCAATCGCTTGACGACAAGTACAAA 1305  
Db 681 AlaLysGlyGlnProValIleTyrTyrGly-----GluGluLeuGlyGlnThrGlyAla 698  
QY 1306 GACTACCTCCCTCGCTCGACCGCTAAGGTGCTGCCGCCATGCTCGATATTGTACGCCGG 1365  
Db 699 AsnAsnTyrProGlnTyrAspAsnArg-----TyrAspPheAlaTrpAsp 713  
QY 1366 CGTATCCTCGCGACAAGCTCCCGATATATTCAAGATGTAATCGACAAGAAATTCAAA 1425  
Db 714 GlnValGluGlyAsnGluIleLeuAlaHisTyrThrLysIleLeuAsnPheArg--Glu 732  
QY 1426 GCGACACGAGAAGTATGACAGACTCGTATTCGACAAGAGTGTGCTTCCTATAGCGAC 1485  
Db 733 GlyTyrSerLysValPheAlaLys-----GlyGluArgThrLeuValGlyGlySerAsp 750  
QY 1486 AAGTTCATGCCATGCTCAAGTCCATCGACAAGCAAGTTCGCCAAGGTATTCGAGAAA 1545  
Db 751 LysAspGlnPheLeuLeuPheSerArgAspTyrGlnAspGlnLysValTyrValGlyLeu 770  
QY 1546 GATCCGCGCAGTAGAGCTTCCAGACGGTA-----ATAGCTGCTGCTGCGCTATT 1596  
Db 771 AsnValAlaGluGlu--SerLysAlaValThrLeuThrValaspSerAlaAspAlaVal 789  
QY 1597 CAGCCCGATGCGATGCGCAATGCC--TATGCCATTCGAGAAGGCAAGCGCTCTTTT 1653  
Db 790 ValThrAspAlaTyrSerGlyThrGluTyrThrAlaThrAlaGlyLys----- 805  
QY 1654 GCCGTTTGGCTGAGATGATACCCCGGAGCTCTCGCGAGGATGCCAATTCACCATG 1713  
Db 806 -----ValAsnLeuThrLeuProGlyLysAlaAspProGlyThrValLeuLeuThr-- 822  
QY 1714 CGTATGAGTACGGCTCCATCAAGGGA----- 1740  
Db 823 --valGluGlyGlyAsnIleThrGlyValAlaLysAspAsnGlyGluValValGlu 841  
QY 1741 -----TATGAACCGCAGGACGGTGCCTGGTACAAAC 1770  
Db 842 LeuValProGluAsnAsnIleArgIleHisTyrLysArgGluAspAsnValTyrLysAsn 861  
QY 1771 TATCATACGACAGCAGCGGCTATTGGAGACGAGATCCTAAGACGGATGAGTTTGC 1820  
Db 862 Tyr-----GlyAlaTrpLeuPaspAspValAlaSerProSerAlaAsnTrpPro 878  
QY 1831 GTACGAGAGATATCCCTCGACCTCTCCGACCAAAACATATGCTGCTATCGCGAG-- 1887  
Db 879 ValGlyAlaThrMetPheGlu-----LysThrAspSerTyrGlyAlaTyrIleAspVal 896  
QY 1888 -----AACGTCAGCTCCATATCGCTTTTCCATGCAACACGACATCAGCGCGGT 1938  
Db 897 ProLeuAlaAspGlyAlaLysAsnIleGlyPheLeuVal--MetAspIleThrAlaGly 915  
QY 1939 AACTCCGCTAGCCCGTATTTCGATAAAGCGCGCTGTGCTGCTTCTGCTTCGATGCC 1998  
Db 916 AspAlaGlyLysAspGlyGlyAspIlePheThrIleSerSerProGlnAlaAsnGlu 935

QY 1999 AACTGGGAAGCTATGAGTGGTGAC-----ATCGAGTTGAAACCC-----GATCTGCAGCGC 2049  
Db 936 IleTrpIleLysGlnGlySerAspLysValTyrThrGluProValaspLeuProAla 955  
QY 2050 ACAATCAGCTGGACATCCGCTACGTTCTCTTCATGATTGAC----- 2091  
Db 956 Asn--ThrValArgIleHisTyrThrArgGluAlaValAspTyrAspAspPheGlyIle 974  
QY 2092 ---AAATGGGGT 2100  
Db 975 TrpAsnTrpGly 978  
RESULT 13  
D97933  
valine-tRNA ligase (EC 6.1.1.9) [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 03-Jun-2002  
C:Accession: D97933  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: D97933  
A:Molecule type: DNA  
A:Status: preliminary  
A:Residues: 1-883 <KOR>  
A:Cross-references: GB:AE007317; PIDN:AAK99296.1; PID:gl5458063; GSPDB:GN00174  
C:Genetics:  
C:Gene: valS  
C:Superfamily: valine-tRNA ligase  
C:Keywords: ligase

Alignment Scores:  
Pred. No.: 0.647 Length: 883  
Score: 133.00 Matches: 104  
Percent Similarity: 33.07% Conservative: 65  
Best Local Similarity: 20.35% Mismatches: 166  
Query Match: 3.48% Indels: 176  
DB: 2 Gaps: 24

US-10-008-355-1 (1-2139) x D97933 (1-883)

QY 574 GATGATTTC-----AAGGACGTTCTGATGTTGTTGTTCTCT 609  
Db 26 AspValPheLysProSerGlyAspGlnLysAlaLysProTyrSerIleValIleProPro 45  
QY 610 CCAGCTCTGTAGTAAGTTCCGAGGCGATACGGACAACCTGGATGTGGCCGCGTCACAGC 669  
Db 46 ProAsnValThrGlyLysLeu-----HisLeu 54  
QY 670 GCGCACTTCAGCGTATTTCGCGTGTATGCCGTCGCGACACCGCGCGCGAATACAGC 729  
Db 55 Gly-----HisAla 57  
QY 730 AAGGACAATAAACCCCTATAAGCCCGTTTACTTCGCTCCGCTATCCATGCAAGGCTACAAG 789  
Db 58 TrpAspThr--ThrLeuGlnAspIleIleIleArgGlnLysArgMetGlnGlyPheAsp 76  
QY 790 GCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCGCTACCTC--ACT 846  
Db 77 -----ThrLeuTrpProGlyMetAspHisAlaGlyIleAlaThr 90  
QY 847 TCTTGGGTGTGAAGATCGTATCGAAGAACAACTCCTCGTATCGAAGTTTCGGGT 906  
Db 91 GlnAlaLysValGluGluArgLeuArgGlyGluGlyIleSerArgTyrAspLeuGlyArg 110  
QY 907 -----ATCAAGCAAGGATCTGGAAGAACCCATGACGCGGATCAGCTAC 954  
Db 111 GluSerPheThrLysValTrpGluTrpLysAsp-----GluTyrAlaThr 126

QY 955 CGTATCAAAATATGCC-----AGCAAGTATGCTCAGAGTCTAACTATTGGAAGAAATTCGATC 1011  
Db 127 ThrileLysGluGlnTrpGlyLysMetGlyLeuSerValAspTyrSerArgGluArgPhe 146  
QY 1012 GGTATGAACCGCGTCTCGCTGCTCTGAGTGTAGTGTCTAGCGTGTAGCGTGGGAGGAAGA 1071  
Db 147 ThrLeuAspGluGlyLeuSer-----LysAlaValArgLys 158  
QY 1072 GCATTCCGACACTGGATCGGTAAG-----AACGCCAAGAGTGTGTCTATGGC 1119  
Db 159 valPheValAspLeuTyrLysGlyTrpIleTyrArgGlyGluPheIleIleAsnTrp 178  
QY 1120 GAT-----GTATTGCTCTCTCGAAAAGGCTTATAAG----- 1152  
Db 179 AspProAlaAlaArgThrAlaLeuSerAspIleGluValIleHisLysAspValGluGly 198  
QY 1153 -----GAAGGACCAAGGCCAACCGTGAGTACT 1182  
Db 199 AlaPheTyrHisMetAsnTyrMetLeuGluAspGlySerArgAlaLeuGluValAlaThr 218  
QY 1183 TATTGAGCGAGAGCTCTTCGGTGGTACCGAGGTGTTGCTTTTGCACAGTTTCCCAAC 1242  
Db 219 ThrArgProGluThrMetPheGlyAspVal----- 228  
QY 1243 GCATTGGCTACAAATCCTGATGCTCATCCCGGTATCCTCAATCGCTTGACGACAAGTAC 1302  
Db 229 AlaIleAlaValAsnProGlu-----AspProArgTyr 239  
QY 1303 AAAGACTACCTCCCTCG-----CTCAGCCGTAAAGTGTGCGCCGCAATG 1347  
Db 240 LysAspLeuIleGlyLysAsnValIleLeuProIleAlaAsnLysLeuIleProIleVal 259  
QY 1348 CTCGAT-----ATTGTACGCCGCGT 1368  
Db 260 GlyAspGluHisAlaAspProGluPheGlyThrGlyValValLysIleThrProAlaHis 279  
QY 1369 ATCCCTCCCGCAAGCTCCCGATATATTCAAGAAATTAATCGACAAGAAATTCAAAAGC 1428  
Db 280 AspProAsnAspPheLeuValGlyGlnArgHisAsnLeuProGlnValAsnIleMetAsn 299  
QY 1429 GACAGGAAGTATGACAGCTTCTGATTCGACAAGAGTGTGGTTCCTATAGCGACAAG 1488  
Db 300 AspAspGlyThrMetAsnGluValPheGluPheSerGlyMet-----AspArg 316  
QY 1489 TTCCATGCTCATGCTCAAGTCCATCGACAAGGAAAGTTTGCACAGGCTATCGACAAGAT 1548  
Db 317 PheGluAlaArg-----LysAlaValAlaLysLeuGluGluIleGly 331  
QY 1549 CCGCAGTAGAGCTTTCCAGAGCGTAAATAGCTGCTGCTCGCGCT----- 1593  
Db 332 AlaLeuValLysIleGluLysArgValHisSerValGlyHisSerGluArgThrGlyVal 351  
QY 1594 -----ATTCAGCGCGAT 1605  
Db 352 ValValGluProArgLeuSerThrGlnTrpPheValLysMetAspGlnLeuAlaLysAsn 371  
QY 1606 GCGATGGCAATGCTTATGCTTATGAGAGGCAAGCGTCTTTTTCGCGGTTGCGT 1665  
Db 372 AlaIleAlaAsnGlnAspThrGluAspLys-----Val 382  
QY 1666 GAGATGTACCCCGACGCTGCTCCCGAGC-----GAT 1698  
Db 383 GluPheTyrProArgPheAsnAspThrPheLeuGlnTrpMetGluAsnValHisAsp 402  
QY 1699 GCCAACTTCACCATCGTATGAGTACAGGCTCCATCAAGGATATGAACCCGACGAGCGT 1758  
Db 403 TrpValIleSerArgGlnLeuTrpTrpGlyHis-----GlnIlePro 416  
QY 1759 GCCTGGTACAACTATCATACGACAGGAGGCGCTATTGGAGAGCAGGATCCTTANGAGC 1818  
Db 417 AlatrptyrAsn-----AlaAspGlyGluMetTyrValGlyGluAlaProGluGly 434

QY 1819 GATGAGTTTCCGCTACAGAGAAATATCTCTCGAC 1851  
Db 435 AspGlyTrpThrGlnAspGluAspValLeuAsp 445  
RESULT 14  
C71528  
probable pts pep phosphotransferase - Chlamydia trachomatis (serotype D, strain UW3/C  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 21-Jan-2000  
C:Accession: C71528  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:9900809; PMID:9784136  
A:Accession: C71528  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-571 <ARN>  
A:Cross-references: GB:AE001306; GB:AE001273; MID:g3328748; PIDN:AAC67931.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
C:Gene: ptsI  
C:Superfamily: phosphotransferase system enzyme I; phosphotransferase system enzyme I  
F:17-565/Domain: phosphotransferase system enzyme I homology <PTI>  
Alignment Scores:  
Pred. No.: 0.628 Length: 571  
Score: 132.00 Matches: 115  
Percent Similarity: 35.37% Conservatives: 88  
Best Local Similarity: 20.03% Mismatches: 186  
Query Match: 3.33% Indels: 185  
DB: 2 Gaps: 24  
US-10-008-355-1 (1-2139) x C71528 (1-571)  
QY 1635 CTTCTCAATGCGATAGCATTTGGCCATCGCATCGCCGCTGAATAGCGGAGCAGCATAT 1576  
Db 22 IleValProGlyIleGlyLeuGlyLysAlaLeu----- 32  
QY 1575 TACGCTCTTGAAAGCTCTACTGCGGATCTTTCTCGATAGCCCTTGCAAACTTTCTCTT 1516  
Db 33 ---LeuLeuGlyLysSerSerLeuArgIleArgGluIleuThrLeuProGlnGluVal 51  
QY 1515 GTCCATGGACTTGACAT-----GGCATGGAACCTTGTGCT 1480  
Db 52 GluHisGluIleSerArgTyrTyrLysAlaLeuLysArgSerArgSerAspLeuAlaAla 71  
QY 1479 ATAAGGAACCACTCTTGTGCGAA-----TACGAAGTCTGCATCTTCTTCTGCTGTC 1429  
Db 72 LeuGluLysGluAlaLysGlyLysGlnGlyTyrGlnGluIleAlaSerIleLeuGlnAla 91  
QY 1428 GCCTTTGAATTTCTGTCGATTACATCTT---GAATATATCGGGAGCTTGTGCGCAGG 1372  
Db 92 HisLeuGluIleIleLysAspProLeuLeuThrGluGluValValLysThrIleArgLys 111  
QY 1371 GATACGCGCGGTACAATATCGACATGCGGCGGAGCACCTTACGCTGCGAGCGGAGG 1312  
Db 112 AspArg-----LysAsnAlaGlu 117  
QY 1311 GTAGTCTTTTACTTTCGTCGTCACCGATTTCAGGATACCGGCATGAGCATCAGG----- 1258  
Db 118 PheValPheSerSerValMetGlyIleGluLysSerLeuCysAlaValGlnLysThr 137  
QY 1257 -----ATTGTGCCAATGCTTGGC 1237  
Db 138 ThrAlaThrArgValAspArgValGlnAspIleHisAspIleSerAsnArgValIleGly 157  
QY 1236 AAACGTGTC-----AAAACGAACCACTCGGTACCCAGGAGCGCTC----- 1192  
Db 158 HisLeuCysCysGlnHisLysSerSerLeuGlyGluPheAspGlnAsnLeuIleValPhe 177  
QY 1192 ----- 1192

Db 178 SerGluGluLeuThrProSerGluAlaAlaAsnAlaAsnProGluTyrIleArgGlyPhe 197  
Qy 1191 -----GCTCAATAGTCTCTCAGCGGTGGCTTGGCTCTCTTCTATA 1147  
Db 198 ValSerLeuGluGlyAlaLysThrSerHisThrAlaIle-----Val 211  
Qy 1146 AGCCTTTTCGAGAGACAATACATCC----- 1117  
Db 212 SerLeuAlaLysAsnIleProTyrValAlaAsnPheThrThrGluLeuTrpAspThrIle 231  
Qy 1116 -----ATAGACAGCACTTTCGCGTCTTACGGATCCAGTCTCGGAATGCTCTTTC----- 1066  
Db 232 LysGluPheSerGlyThrLeuValLeuIleAsnGlyAspLysGlyGluIleThrPheAsn 251  
Qy 1065 -----CTCGCAGCTTACGACCTATCAGCTCAAG-----ACGAGC 1030  
Db 252 ProGlnLeuSerThrIleGlnThrTyrTyrArgLysGlnAlaSerValSerValThrVal 271  
Qy 1029 GAGACCGCGGTTCATACGATCGAATCTTCCAAATAGTTCAGTCTGAGCATCTTGTCT 970  
Db 272 ProValGlnValGlnThrGlyLysAsnLeuProLeuIleSerLeuSerAla----- 288  
Qy 969 GGCATATTTGATACGGTAGCTGATCTGCGCTCATGCTCTCCCTCCAGATCCCTTCTT 910  
Db 289 -----GlnIleValSerThrGluGluLeuProMetIle 299  
Qy 909 G-----ATACCGGAACCTTCGATACGAGGATGTTCTCGTTTCGATACGATCTTC 859  
Db 299 eGluArgGluSerProGlyThrSerVal--GlyLeuPhe----- 311  
Qy 858 CACACCCCAAGAAGTGAGGTAGCTGATCTGCGCTCATGCTCTCCCTCCAGATCCCTTCT 799  
Db 312 -----ArgSerGluPheMetAlaPhe 318  
Qy 798 GTCGTACCCCTGTAGCTTCGATCGATACGCGACGCAAGTAACGGCTTATAGGTTT 739  
Db 318 eSerLeuGlyArgLeuProCysValGlu-GluGlnAlaAspGlnTyrAlaGlnLeuValG 338  
Qy 738 AT----- 736  
Db 338 InPheGlnCysSerAspIleHisValLeuArgLeuPheAspPheGlyGluAspLysGluC 358  
Qy 735 GTCCTTGTCTGATTGGCGCGCGGTTCGCGCACCGGCTACACCGGAATACGCTGAA 676  
Db 358 ysProCys-----IleSerSerHisArgSerValArgTrpLeu-LeuGlu 373  
Qy 675 GTCGCCCTGTGACGCGCCACAT--CCAGTTGCTGCTATGCTCCGCTCCGAACTTACCTAC 619  
Db 374 GlnGluLysValLeuLysGluGlnLeuGlnAlaIleValSer----- 389  
Qy 618 AGAGTGGGAGGAGCAATACCATACGAACTCTTGAATACATCGTAGACGATGAGGAA 559  
Db 390 ArgIleGlyArgLeuLysVal-----LeuIleProGlyValIleAspAlaSerGlu 406  
Qy 558 GTATTCGTGTGTAATAGAAAGCTCTACGATCGAGTGTGTTCTGCTGCTGCTATTTTC 499  
Db 407 IleAlaLeuVal-----LysArgLeuPheGlnGluIleArgLeuLeuLysGlyIle 424  
Qy 498 TTTTTCGCCAGTCTTGGCATACCTCTGAGCTTGGCAGACGCTCCATCTGCTCAGT 439  
Db 425 SerGluAsnIleLeuTrpGlySerMetIleGluIleProSerAlaValTrpMetIleGlu 444  
Qy 438 CATACCTTGGCTGCTCTCTTCTGCTGCTGCTTACCTTACGATCTTGGCAGATACCTT 379  
Db 445 GluIleLeuGlnLeuSerPheValAlaLeuGlyThrAsnAspLeuAlaGlnTyrThr 464  
Qy 378 CACGGAAGACCCGGAATCGGAAGCTCTCACCATCGTGGAGAAACCAACCATCGCG 319  
Db 465 LeuGlyThrSerArgGluArgSerLeuLeuGlyGluArgSerArgValProHisProSer 484  
Qy 318 CAGATAGCTGATCCACCGCTGCTTG-----CCTCTGGATACGACCGTATCCGCGAGT 265  
Db 485 ValIleArgMetIleHisValValGluGlnAlaLysGlnLysAsnValProValSer 504

Qy 264 GTGGTGTGTAATA-----GATCAGGCCCTGATC 238  
Db 505 ValCysGlyGluMetAlaGlyAspProAlaLeuLeu 516

## RESULT 15

PC4395  
mucin 3 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 10-Nov-1997 #sequence\_revision 10-Nov-1997 #text\_change 08-Oct-1999  
C:Accession: PC4395  
R:Van Klinken, B.J.W.; Van Dijken, T.C.; Oussoren, E.; Bueller, H.A.; Dekker, J.; Ein  
Biochem. Biophys. Res. Commun. 238, 143-148, 1997  
A:Title: Molecular cloning of human MUC3 cDNA reveals a novel 59 amino acid tandem re  
A:Reference number: PC4395; MUID:97445141; PMID:9299468  
A:Accession: PC4395  
A:Molecule type: mRNA  
A:Residues: 1-648 <VAN>  
A:Cross-references: DDBJ:AF016692; NID:g2454614; PIDN:AAB71685.1; PID:g2454615  
A:Experimental source: intestine  
F:1-59,60-118,119-177,178-236,237-295/Region: repeat  
F:296-565/Region: semi-unique #status predicted  
F:566-582,583-599,600-616,617-633,634-647/Region: repeat

Alignment Scores:  
Pred. No.: 0.776 Length: 648  
Score: 131.00 Matches: 140  
Percent Similarity: 34.14% Conservative: 71  
Best Local Similarity: 22.65% Mismatches: 173  
Query Match: 3.43% Indels: 234  
DB: 2 Gaps: 34

US-10-008-355-1 (1-2139) x PC4395 (1-648)

Qy 536 CTTTCTATTCCAAACAAGATACCTCTCATCTCTAGATG-----TATTCAAGACGTTTC 592  
Db 104 MetSerThrProSerGluGlySerSerSerLeuThrThrMetLeuLeuSerSerThrThr 123  
Qy 593 GTATGGTATTGTCTCTCCCTCCAGCTCTGTAGTAACTTCGGAGCGGATA----- 640  
Db 124 Val-----ThrSerSerGluAlaSerThrProSerThr 134  
Qy 641 -----CGACAACCTGGATGCGCGCTCATCGGCGACTTCACGCGTATTTC 688  
Db 135 ProSerValAspArgSerThr-----ProValThrThrSerThrGlnSerAsnSer 151  
Qy 589 GCG-----TGATGCGGTGGCGACACCGCGCGCCGCAATACA 727  
Db 152 ThrProThrProProGluValIleThrLeuProMetSerThrProSerGluValSerThr 171  
Qy 728 GCAAGGACAATAAACCTATATAAGCCGTTTACTTTCGCTGCGCTATCCATCGCAAGCTACA 787  
Db 172 ProLeuThrIleMetProValSer-----ThrThrSerValThrIleSerGluAlaGly 189  
Qy 788 AGGCTGACGACTATGCCATCACCATCGGTTTCCGCGGCGTACGCGTACCTCA--- 844  
Db 190 ThrAlaSerThrLeuPro-----ValAspThrSerThrProValIleThrSerThr 206  
Qy 845 -----CTTCTGGGTGTGGAAGATC----- 865  
Db 207 GlnValSerSerProValThrProGluGlyThrThrMetProIleThrThrLeuSer 226  
Qy 866 -----GTATCGAAACGAGAACA----- 883  
Db 227 GluGlySerIleProIleThrIleMetHisValSerThrThrArgValThrSerSerGlu 246  
Qy 884 -----ATCCTCGTATCGAAGTTC 901  
Db 247 GlySerThrLeuSerThrProSerValValThrSerThrProValThrThrSerThrGlu 266  
Qy 902 GCGGTATCAAGCAAGGATCTGGAAGAACCCATCAGCGGAGATCAGGCTACCCGCTATCA 961  
Db 267 AlaIleSerSerSerAlaThr-----LeuAspSerThr 977







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:15:31 ; Search time 36.5 Seconds  
(without alignments)  
4861.253 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 3820

Sequence: 1 atgcaatgaattaaaaaag.....aagagctgaagtgtgatctaa 2139

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=Cgn2.1/USPTO.spool/US10008355/runat\_16052003.110337.9320/app.query.fasta\_1.2311  
-DB=SwissProt\_40 -QFMT=fastan -SURFLX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355.ecgn.1.1.47.erunat\_16052003.110337.9320 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134.5	4.0	1609	1 FIG2_YEAST	P25653 saccharomyc
2	145	3.8	1367	1 AMYH_YEAST	P08640 saccharomyc
3	143.5	3.8	776	1 YLA4_CAEEL	Q05036 caenorhabdi
4	143.5	3.8	1802	1 HKR1_YEAST	P41809 saccharomyc
5	139	3.6	4543	1 LRPL_CHICK	P98157 gallus gall
6	137.5	3.6	1140	1 YW96_YEAST	Q04893 saccharomyc
c 7	132	3.3	571	1 PTL_CHLTR	O84340 chlamydia t
8	127.5	3.3	725	1 AGAL_YEAST	P32323 saccharomyc
9	126.5	3.3	3178	1 YS89_CAEEL	Q09624 caenorhabdi
10	122	3.2	1104	1 COLA_CLOPE	P43153 clostridium
11	121	3.2	1628	1 NAGH_CLOPE	P26831 clostridium
12	118	3.1	698	1 EFG_VIECH	Q9Kuz7 vibrio chol
13	117.5	3.1	1024	1 HLYA_ECOLI	P08715 escherichia
14	117	3.1	1147	1 TEAL_SCHPO	P87061 schizosacch
15	116.5	3.0	1169	1 YK82_YEAST	P36170 saccharomyc
c 16	116	2.9	495	1 MURE_XYLFA	Q9pf85 xylella fas
17	116	3.0	652	1 TRGA_VIECH	P27772 vibrio chol
18	116	3.0	952	1 TOP1_STRCO	Q9x909 streptomyce

19	114	3.0	620	1	DNAK_PORPU	P30723 porphyra pu
20	114	3.0	734	1	GLGB_AGRTO	P52979 agrobacteri
21	114	3.0	747	1	VIVC_BPT7	P03725 bacterioph
22	113	3.0	507	1	YG46_YEAST	P53301 saccharomyc
23	113	3.0	1047	1	RIRL_CHLMU	Q93193 chlamydia m
24	113	3.0	1161	1	DAN4_YEAST	P47179 saccharomyc
25	113	3.0	1273	1	WEB1_YEAST	P38968 saccharomyc
26	113	3.0	4687	1	PLE1_RAT	P30427 rattus norv
27	112.5	2.9	657	1	CN16_HAEIN	P44764 haemophilus
28	112	2.9	556	1	EST2_CAEEL	Q07085 caenorhabdi
29	112	2.9	1013	1	PPOL_HUMAN	P09874 homo sapien
30	112	2.9	1504	1	DPOZ_YEAST	P14284 saccharomyc
31	111.5	2.9	878	1	ACON_RICPR	Q9zcf4 rickettsia
32	111.5	2.9	922	1	YKL6_CAEEL	P42173 caenorhabdi
33	111.5	2.9	1023	1	HLV1_ECOLI	P09983 escherichia
34	111.5	2.9	2410	1	MOK1_SCHPO	Q9ust8 schizosacch
35	110.5	2.9	498	1	FLUD_BAGSU	P39738 bacillus su
36	110.5	2.9	1039	1	STL_METJA	Q58357 methanococc
37	109.5	2.9	768	1	PARC_NEIGO	P48374 neisseria g
38	109	2.9	681	1	MP10_HUMAN	O00566 homo sapien
39	109	2.9	3672	1	LML2_CAEEL	Q21313 caenorhabdi
40	108.5	2.8	507	1	YY42_CAEEL	Q18416 caenorhabdi
41	108.5	2.8	1041	1	DD16_HUMAN	O60231 homo sapien
42	108	2.8	638	1	NBC2_PIG	Q30333 sus scrofa
43	108	2.8	1341	1	RPAL_METJA	Q38445 methanococc
44	107.5	2.8	636	1	YNR6_YEAST	P53882 saccharomyc
45	107.5	2.8	666	1	UVRB_CLOAB	Q971q2 clostridium

#### ALIGNMENTS

##### RESULT 1

FIG2\_YEAST STANDARD; PRT; 1609 AA.  
AC P25653;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Factor induced gene 2.  
GN FIG2 OR YCR089W OR YCR89W OR YCR1102.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92397594; PubMed=1523889;  
RA Wilson C., Grisanti P., Frontali L.;  
RT "The complete sequence of a 6146 bp fragment of Saccharomycetes  
cerevisiae chromosome III contains two new open reading frames.";  
RL Yeast 8:569-575(1992).  
CC -!- FUNCTION: REQUIRED FOR EFFICIENT MATING.  
CC -!- INDUCTION: BY MATING PHEROMONES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL: X59720; CAA42254.1; -  
DR PIR: S19504; S19504.  
DR PIR: S25345; S25345.  
DR SGD: S0000685; FIG2.  
SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;

##### Alignment Scores:

Pred. No.: 0.0162 Length: 1609  
Score: 154.50 Matches: 153  
Percent Similarity: 35.18% Conservative: 101  
Best Local Similarity: 21.19% Mismatches: 254

Query Match: 4.04% Indels: 214  
DB: 1 Gaps: 28

US-10-008-355-1 (1-2139) x FIG2\_YEAST (1-1609)

QY 74 ACAAGGATGCTGCTCCTCAAGAACTCAATCAGGAGAACTCGATCGCAATGCGTGAGC 133  
DB 788 ThrGluValCysSerHisSer-----GluCysThrPro 798  
QY 134 TCGGCTTTACGCTCCCGTGGATTCGCTACAGCTTTTCGACAAACCCGTCATCCCAATG 193  
DB 799 ThrValIleThrSer-----ValThrAlaThrSerThr-----Ile 811  
QY 194 CCGTGGTATTCTTCGGTGGCGGATGATACCGGTATCACAGTTCGATCCGAGGCGCTGATCT 253  
DB 812 ProLeuLeuSerThr-----Ser 817  
QY 254 TTACCAACCACTGGGATGACGCTGATCCAGAGCCAAA-----GCAGGTGGATC 307  
DB 818 SerSerThrValLeuSerSerThrValSerGluGlyAlaLysAsnProAlaAlaSerGlu 837  
QY 308 ACGACTATCTCGCGGATGGTTTCTTCGACGATGGGTGAGGAGCTTCGATTCGG 367  
DB 838 ValThrIleAsnThrGlnValSerAlaThrSerGluAlaThrSerThr-----853  
QY 368 GTCTTTCGCGTGAAGTATCTCGCGCAAGATCGTGAAGGTAAACGGCAAGGTAGAGACAGC 427  
DB 854 -----Ser 854  
QY 428 TCAAGGGTATCACTGACGAGATGGAGCGTCTGGCAAGCTCAGGAGGTATGCCAAGAAC 487  
DB 855 ThrGlnValSerAlaThr-----860  
QY 488 TGGCCAAAAGAAATCCAGCAGAGAACCAACTCGCATCTGATAGAGCTTTCTATTCCA 547  
DB 861 -----SerAlaThrAlaThrAlaSerGluSerSerThrThrSerGlnVal 875  
QY 548 ACAAGGAATACTTCTCATCTGCTACGATGATTCAGAGGCTTC-----GTATGGTAT 601  
DB 876 SerThrAlaSerGluThrIleSerThrLeuGlyThrGlnAsnPheThrThrThrGlySer 895  
QY 602 TTGCTCTCCAGCTCTGTAGTGTAGTTCGGAGGCGATACGGACAACTGGATGGCGCGC 661  
DB 896 LeuLeuPheProAlaLeu-----SerThrGluMetIleAsnThrThrValValSerArg 913  
QY 662 GTCACACGGCGACTTCA-----GCGTATTCGCGGTGATCGCGTGGCGGACA 709  
DB 914 LysThrLeuIleSerThrGluValCysSerHisSerLysCysValProThrValIle 933  
QY 710 ACCGGCGCGCGGAATACAGCAAGGACAAATACCCTATAAGCCCGTTTACTTCGTGCGG 769  
DB 934 ThrGluValValThrSerLysGlyThr-----ProSerAsnGlyHisSerSerGlnThr 951  
QY 770 TATCCATCGAAGCTACAAAGCTGAGGCTAGGACTATGCCATGACCATCGTTTCCCGGCAGTA 829  
DB 952 LeuGlnThrGluAlaValGluValThrLeuSerSerHisGlnThrValThrMetSerThr 971  
QY 830 CGGATCGCTACCTACTTCTGGGTGGTGAAGATCGTATCCGAAACACGAGA-----880  
DB 972 GluValCysSerAsnSerIleCysThrProThrValIleThrSerValGlnMetArgSer 991  
QY 881 -----ACAATCTCTGATTCGAAGTTCGGGTATCAAGCAAGCGATCTCGAAGGAG 931  
DB 992 ThrProPheProTyThrLeuThrSerSerThrSerSerSerSerLeuAlaSerThrLysLys 1011  
QY 932 CCATGAGCCAGATCAGGCTACCGGTATCAATATGCCAGCAAGTATGCTCAGAGTGCTA 991  
DB 1012 -----SerSerLeuGluAlaSer-----1017  
QY 992 ACTATTGGAAGAAATTCGATCGGTATGAACCGCGGTCTCGCTGTGACGTGATAGGTC 1051  
DB 1017 -----1017

QY 1052 GTRAGCGTCCGAGGAAGAGCATTCCGACACTGGATCCGTAAGACGCGCAAGAGTCTG 1111  
DB 1018 -----SerGluMetSerThrPheSerValSerThr-----GlnSerLeu 1030  
QY 1112 TCTATGCGCATGATTGCTCTCTCTCGAAAGGCTTATAAGGAAGGAGCAAGCCCAAC 1171  
DB 1031 ProLeuAlaPheThrCys-----SerGluLysArgSerThr 1042  
QY 1172 GTGAGATGACTTATTTGAGCGAGACGCTCTCGGTGGTACCGAGGTGGTTCGTTTCCAC 1231  
DB 1043 -----ThrSerValSerGlnTrp-----1048  
QY 1232 AGTTTGCACAGCATTCGCTACAAATCCTGATGCTCATGCGGTATCTCTCAAAATCGCTG 1291  
DB 1049 -----SerAsnThrValLeuThrAspThrIleMetSerSerSerAsnValIle 1065  
QY 1292 ACGACAAGTACAAAG-----ACTACCTCCCTCGCTCGACCGCTAAGGTGCTGCCCG 1342  
DB 1066 SerThrAsnGluLysProSerSerThrThrSerProTyThrAsn-----1079  
QY 1343 CCATGCTCGATATTGTACCGCGGCTATCCCTGCCGACAAAGTATCCCGATATATTTCAAGA 1402  
DB 1080 -----PheSerSerGlyTy:SerLeuProSerSerSerThrProSerGlnTy 1095  
QY 1403 ATGTAATCGACAGAAATTCAAAGCGCACACCAAGAGATATGCAGACTTCGTATTCCGACA 1462  
DB 1096 SerLeuSerThr-----AlaThrThrIleAsnGlyIleLysThrValTyThrThr 1113  
QY 1463 AGATGTGGTTCCTTATAGCGCAAGTTCATGCGCATGCTCAAGTCCATGCCAGCAAGGAAA 1522  
DB 1114 -----TrpCysProLeuAlaGluLysSerThrValAlaAlaSerSerGlnSerSerArg 1131  
QY 1523 AGTTTGCACAGGCTA-----TCGAGAAAGATCCGCGAGTAGAGCTTTTCCAAAGACGCTAA 1576  
DB 1132 SerValAspArgPheValSerSerSerLys--ProSerSerSerLeuSerGlnThrSerI 1151  
QY 1577 TAGCTGCTGCTCGGCTATTAGCGCGGATGCGCAATGCCATGCCATGCCATGCCATGAGA 1636  
DB 1151 le-----GlnTyThrLeuSerThrA 1158  
QY 1637 GCAAGCGCTTTTCTTTCGCGGTTTTCGCGTGGATGATACCC-----1677  
DB 1158 laThrThrThr---IleSerGlyLeuLysThrValTyThrThrTrpCysProLeuThrS 1177  
QY 1678 -----GGAGCTGCTCTGCGGAGGATGCCAACTTCCACATGGTATGAGCT 1723  
DB 1177 erLysSerThrLeuGlyAlaThrThrGlnThrSerSerThrAlaLysValArgIleThr- 1196  
QY 1724 ACGGCTCCATCAAGGATATGAACCGCAGGAGGCTGCTGCTACAACTATCATACGACAG 1783  
DB 1197 SerAlaSerSerAlaThrSerThrSerIleSerLeuSerThrSerThrGluSerGluSer 1216  
QY 1784 GCAAGGGGCTATTGGAGAGCAGGATCCTAAGAGCGATGAGTTTCCCGTACAGGAGAATA 1843  
DB 1217 SerSerGlyTyThrLeuSerLysGlyValCysSerGlyThrGluCysThrGlnAspValPro 1236  
QY 1844 TCCTCGACCTCTTCCGACCAAAACTATGTCGCTATGCCGAGAGACGTCAGCTCCATA 1903  
DB 1237 ThrGlnSerSerSerProAlaSerThrLeuAlaTySerProSerValSerThrSerSer 1256  
QY 1904 TCGCTTCTCTATCAACCAACGACATCACGGGGGTAAC-----CCG 1945  
DB 1257 SerSerPheSerThrThrThrAlaSerThrLeuThrSerThrHisThrSerValPro 1276  
QY 1946 GTAGCCCGTATTTCGATGAAGACCGCGCTCTGATCGGTCTTCTTTCGATGCAACTGG 2005  
DB 1277 LeuLeuProSerSerSer-----SerIleSerAlaSerSerProSerSer 1291  
QY 2006 AAGCTATGAGTGGTGGATTCGAAACCGGATCTCGCAACCGCAATCATCCGCTGGACA 2065  
DB 1292 ThrSerLeuLeuSerThrSerLeuProSerProAlaPheThrSerSerThrLeuProThr 1311  
QY 2066 TCCGCTACTCTCTTCATGATTGACAAATGGGTGCGTCCCGCTCTCATCCCAAGAGC 2125

```

Db 1312 AlaThrAlaValSerSerThrPheIleAlaSerSerLeuProLeuSerSerLysSer 1331
      ...      |||||      |||      |||||      |||||      |||||
RESULT 2
AMYH_YEAST
ID AMYH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
GN STAL OR STA2 OR MAL5 OR YI019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Iye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAL.";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
RT from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; Z38061; CAA86176.1; -
DR EMBL; M16164; AAA35014.1; -
DR EMBL; M16165; AAA35015.1; -
DR EMBL; X13857; CAA32069.1; -
DR PIR; B26877; B26877.
DR PIR; A26877; A26877.
DR PIR; S48478; S48478.
DR SGD; S0001458; MUC1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCOSYLASE S1/S2.
FT DOMAIN 210 1367 SER/THR-RICH.
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

```

```

Alignment Scores:
Pred. No.: 0.0657 Length: 1367
Score: 145.00 Matches: 160
Percent Similarity: 33.69% Conservative: 95
Best Local Similarity: 21.14% Mismatches: 354
Query Match: 3.80% Indels: 148
DB: 1 Gaps: 25

US-10-008-355-1 (1-2139) x AMYH_YEAST (1-1367)
QY 65 CCAAGCGGACAAAGGATGCTGCTCTCAACGAACATCAATCAGGAGAAATCTGGATCGAA 124
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 355 ProValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSer 374
QY 125 TGGCGTGAAGTCGGCTTTTACGCTCCCGTTGGATTCGCTCTACAGTTTCGACAAAGCGTCCA 184
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 375 ThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaPro 394
QY 185 TTGCCAATGCCGTGGTATATCTTCGGTG-----GCCGATGTACGGGTATCAGTGTCCG 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 395 ValProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThr 414
QY 239 ATCAGGCGCTGATCTTTTACCAACACACACACGATGCGGATACGCTGTCTCCAGGACCAAGCA 298
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 434
QY 299 CGGTGGATACGACGATATCTCGCGGATGTTTCGTTCTCGCAGGATGGTGAGGAGCTTC 358
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 ThrSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSer 454
QY 359 CGATTCGGGTC---TTTCGGTGAAGTATCTGCGCAAGATCGTGGAAGGTAAAGGACAAAGG 415
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 SerAlaProValProThrProSerSerSerThrThrGluSer----- 468
QY 416 TAGAAGGACAGCTCAAGGGTATCCTGACGAGATGGAGCGTCTCGCAAGCTCAGGAGG 475
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 469 ---SerSerAlaProValThrSerSerThrThrGluSerSerSerAlaProValProThr 487
QY 476 TATGCCAAGAACTGGCCAAAAGAAATCGACGAGAGAACCACTCTGCATCTGCATGAGC 535
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 488 ProSerSerThrThrGluSerSerSerAlaProValThrSerSerThrThrGluSer 507
QY 536 CTTTCTATTCCACAGGAATATCTCTCATGCTGTACGATGTATTCAAGGACGCTTCGTA 595
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 508 SerSerAlaProValProThrProSerSerSerThrThrGluSerSerAla----- 525
QY 596 TGGTATTGCTCTCCAGCTCTCTAGTAGTTCGGAGGCGATACGACAACTGGATGT 655
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 526 -----ProAlaProThrProSerSerSerThrThrGluSerSerSerAla 540
QY 656 GGCCGGCTACACGGGGGACTTCAGCGTATTTCGCGGTGTATGCGGTCGCGACAAACCGGC 715
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 ProValThrSerSerThrThrGluSerSerSerAla-----ProValProThrProSer 558
QY 716 CGGCGGAATACAGCAAGGACAAATAACCCCTATAAGCCCGTTTACT-----TCG 763
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 SerSerThrThrGluSerSerSerThrProValThrSerSerThrThrGluSerSerSer 578
QY 764 CTGCGGTATCCA-----TGCAAGGCTACAAAGGCTGACGACTATGCCATGACCATCG 814
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 579 AlaProValProThrProSerSerSerThrThrGluSerSerSerAlaProValProThr 598
QY 815 GTTTCCCGGCGAGTACGATCGCTACCTCCTCTTCTTGGGTGTGGAAGATCGTATCGAAA 874
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 ProSerSerThrThrGluSerSerSerSerAlaProAlaProThrProSerSerSerThr 618
QY 875 ACGAGAACAAATCTCGTATCAAGTTCGGGTATCAAGCAAGGATCTGGAAGGAAGCA 934
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 619 ThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerSer----- 635
QY 935 TGACCGGATACGAGCTACCGGTATCAATATGCCACCAAGTATGCTCAGAGTCTCAACT 994
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 636 ---AlaProValProThrProSerSerSer-----ThrThr 646

```

```
Qy 995 ATTGGAAGATTTCGATCGGTATGAACCGCGGTCTCGCTCGTTCGACGTGATGTCGTA 1054
Db 647 GluSerSerAlaProValProThrProSerSerSerThrThrGluSerSerAla 666
Qy 1055 AGCGTGGCAGAAAGACATTCGAGACTGATGC-----GTAAGAAGCGCA 1102
Db 667 ProValProThrProSerSerSerThrThrGluSerSerAlaProValThrSerSer 686
Qy 1103 AGATGCTGTCTATGGCGATGATTGTCTTCTCGAAAGAGCTTATAGGAGAGCA 1162
Db 687 ThrThrGluSerSerAlaProValThrSerSerThrThrGluSerSerAlaPro 706
Qy 1163 AGGCCAACCGTACGATGACTTATTGAGCGAGAGCTCTCGGTGGTACCGAGGTGGTTC 1222
Db 707 ValProThrProSerSerSerThrThrGluSerSerAlaProValPro----- 723
Qy 1223 GTTTTGCACAGTTGCCAACGCGATTGGCTAGCAAAATCTGATGCTCATCGCGTATCTCA 1282
Db 724 ThrProSerSerSerThrThrGluSerSerAlaProValProThrProSerSerSer 743
Qy 1283 AATCGCTTGACGACAGTACAAAGACTACCTCCCTCCCTCGACCGTAAAGTGTGCGCCG 1342
Db 744 ThrThrGluSerSerAlaProValThrSerSerThrThrGluSerSerAlaPro 763
Qy 1343 -----CCATGCTCGATATTGTACCGCGCGGTATCCCTCGCGACAAAGTCCCGGATA 1393
Db 764 ValProThrProSerSerSerThrThrGluSerSerAlaProValProThrProSerSer 783
Qy 1394 TATTCAAGAATGTATCG-----ACAGAAATTTCAAAGCGGACACGA 1435
Db 784 SerSerThrThrGluSerSerAlaProValProThrProSerSerSerThrThrGlu 803
Qy 1436 AGAGTATG-----CAGACTTCGTATTCGACAGAGTGTGTTCTTCTTATACG 1483
Db 804 SerSerValAlaProValProThrProSerSerSerSerSerSerSerSerSerSer 818
Qy 1484 ACAAGTTCATGCCATGCTCAAGTCCATGACGACAAAGAAAGTTTGCCAAAGCTATCGAGA 1543
Db 819 ThrSerSerAlaProSerSerThrThrProPheSerSerSerThr----- 832
Qy 1544 AAGATCCGCGAGTAGAGCTTTCCAAAGACGTAATAGCTGTGCTCGCCCTATTCAGGCCG 1603
Db 833 -----GluSer 834
Qy 1604 ATCGATGCGCAATGCCATGCTATGAGAGGCGCAAGCGTCTTCTTCTCGCGTTTGC 1663
Db 835 SerSerValProValProThrProSerSerSerThrThrGluSerSerAlaProVal 854
Qy 1664 GTGAGATGACCCGCGAGTGTCTGCGCGAGCGATGCCAAT-----TCACCATGC 1714
Db 855 SerSerSerThrThrGluSerSerValAlaProValProThrProSerSerSerSerSer 874
Qy 1715 GTATGAGTACGGTCCATCAAGGATATGAACCGCAGGAGCGTGCCT----- 1762
Db 875 IleThrSerAlaProSer-----SerIleProPheSerSerThr 888
Qy 1763 -----GGTACACTATCATACGACAGGCAAGGCGTATTGGAGAGC 1804
Db 889 ThrGluSerPheSerThrGlyThrThrValThrProSerSerSerSerSerSerSer 908
Qy 1805 AGGATCTCTAAGAGCGATGAGT-----TTGCCCTACAGAGAGATA 1843
Db 909 GlnThrGluThrSerValSerSerThrThrGluThrThrIleValProThrLysThrThr 928
Qy 1844 TCTCGACCTCTTCGGCACCAAAACTATGTC----- 1876
Db 929 ThrSerValThrThrProSerThrThrIleThrThrThrValCysSerThrGlyThr 948
Qy 1877 -----GCTATGCCGAGACGCGTACGCTCCATATCGCTT 1909
Db 949 AsnSerAlaGlyGluThrThrSerGlyCysSerProLysThrValThrThr----- 965
```

```
Qy 1910 TCCTATCGAACAAGCATCACGGCGGTAACTCCGCTAGCCCGCTATTTCGATAAGAAGC 1969
Db 966 ThrValProThrThrThrThrThrSerValThrThrSerSerThrThrThrThrThr 985
Qy 1970 CGCGTCTGATCGGTCTTCCTTCGATGCGCAACTGGGAAGCTATGAGTGGTGACATCGAGT 2029
Db 986 ThrVal-----CysSerThrGlyThrAsnSerAlaGlyGluThrThrSerGly 1001
Qy 2030 TCGAACCG-----ATCTGCAGCGCAACATCAGCGTGGACATCC 2068
Db 1002 CysSerProLysThrIleThrThrThrValProCysSerThrSer-----ProSer 1018
Qy 2069 GCTACGTTCTTCTTCATGATGACAAATGGGTGCTAGTCCCGCTCTCATCC 2119
Db 1019 GluThrAlaSerGluSerThrThrThrSerProThrThrProValThrThr 1035

RESULT 3
YLA4_CAEEL
ID YLA4_CAEEL STANDARD: PRT; 776 AA.
AC Q05036;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 86.9 kDa protein C30C11.4 in chromosome III.
GN C30C11.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Cooper J., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Durbin R., Favello A., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisner N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith K., Sonnhammer E., Staden K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans".
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY. STRONG,
CC TO YEAST MS13.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L09634; AAA27967.1; -.
DR PIR; S44784; S44784.
DR WormPep; C30C11.4; CE00103.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR ProDom; PD000089; Hsp70; 1.
DR PROSITE; PS00297; HSP70_1; FALSE_NEG.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR KW Hypothetical protein; ATP-binding; Multigene family.
SQ SEQUENCE 776 AA; 86896 MW; 20FA975FB649FA9C CRC64;

Alignment Scores: 0.0687 Length: 776
Pred. No.: 143.50 Matches: 118
Score:
```

Percent Similarity:	37.48%
Best Local Similarity:	21.57%
Query Match:	3.76%
DB:	1

US-10-008-355-1 (1-2139) x YLA4\_CAEEL (1-776)

Qy	113	ATCTGATCGAATCGGTGACGTGGCTTTACGCTCCCGTTGGATTCCGCTCTACAGTTTCG	172
Db	236	ileTrpPheAsp-----AlaLeuileArgGluHisPheArgLeuGlyAsnGluPheLys	251
Qy	173	ACAAGCGCTCCATTG-----CCAATGCCGTGGTTATCTTCGGTGGCGGAT	217
Db	252	ThrLysTrpGlyLeuAspAlaalaThrSerProArgProThrPleuArgLeuLeuAspGlu	271
Qy	218	GTACCGGTATCACAGHGTCCGATCAGGCGCTGATCTTTACCAACACACACTGCGGATPACG	277
Db	272	CysGluArgValLysLysGlnMetSerAlaAsnGlnThrProIle-ProLeuAsnIleG1	291
Qy	278	GTGCTATCCAGACCAACACGGTGGATCAGCACTATCTCGCGGATGGTTTCTGTTCTC	337
Db	291	uCysPheMetGluAspLysAspValThrGlyLysMetGln-ArgGlnCluPhe-----	308
Qy	338	GCACGATGGGTGAGGAGCTTCGGATTCGG-----	366
Db	309	-----GluAspLeuAlaAlaProIlePheAsnArgIleLysGlnValLeuIleA	325
Qy	367	-----GGTCCTTCGGTGAAGTATCTCGCGCAAGATCGTGAAGTACGACACA	412
Db	325	snLeuPheAlaAspGlyValSerIleLys-----ProGluGluLeuAspG	340
Qy	413	AGGTAGAAGACAGCTCAAGGGTATCTACTGACGAGATGGAGCTTCTATGCCAACCAAT	472
Db	340	luIleGlu-----IleValGlyGlySerSerArgIleProMetIleArgG	355
Qy	473	AGGTATGCCAAGACTGGCCAAAAGAAG-----AATGCAGACAG-----	513
Db	355	luIleValLysAspLeuPheGlyLysGluProLysThrMetAsnGlnAspGluAlav	375
Qy	514	-----AACCACACTCTGCATCGTAGAGCCTTCTATGCCAACCAACAAAT	556
Db	375	alAlaArgGlyAlaAlaMetGlnCysAlaIleLeuSerProThrPheArgValArgGluP	395
Qy	557	ACTTCCTCATCTGTACGATGATTCACGAGC-----GTTCGTATGTGTAT	601
Db	395	he-----AlaIleLysAspThrGlnProTyrArgIleArgLeuSerT	409
Qy	602	TTGCTCTCCACAGCTCTGTAGTGAAGTTCGGAGGCGGATACGCAACACTGGATGTGGCCGC	661
Db	409	rp-----AsnSerThrGlyGluAsnGlyCylGluAsnAspValPheSer---ProA	425
Qy	662	GTACACAGGGCGACTTCAGC-----GTATTCGGCGTGTATCGCGGTGCCGACACGGCG	715
Db	425	rgAspGluValProPheSerLysLeuValSerLeuLeuArgSerGlyProPheAsnValG	445
Qy	716	CGGCGGAATACACAGGACAAT-----AAACCCCTATAAGCCGTTTACTTCCTCCCGGTAT	772
Db	445	luAlaHisIstyrAlaGlnProAsnValValProHisAsnGlnValHisIleGlySerTrpL	465
Qy	773	CCATCGAAGGCTACAAGCTGACGACTATGCCATCACCATCGTTTCCCGGGCAAGTACGG	832
Db	465	ysValAsnGlyAlaArg-----ProGlyAlaAsp-----	474
Qy	833	ATCGCTACCTCACTTCTTGGGTGTGGGAAGATCGTATCGAAAACGAGAACAACTCCTCGTA	892
Db	475	-----GlyGlyAsnGlnLysVal-----LysV	482
Qy	893	TCGAAGTTCCGGTATCAACGAACGATCTGGAAGAGCCATGATGAGCGCAGATCAGGCTA	952
Db	482	alLysValArgValAsnProAspGlyIlePheThrIleAla-----	495
Qy	953	CCCGTATCAAATATGCCAGCAAGTATGCTCAGAGTGCTCAACTATTGGAAGAATTTCGATCG	1012

Db	496	-----SerAlaThrMetTyrGluProArgIleVal	505
Qy	1013	GTATGAACCGCGTCTCGCTGCTTACGATGATAGTTCGTAAGCGTCGCCGAGGAAAGAG	1072
Db	505	alGluGluValProAlaGluAlaMetGluValAspLysThrGluAlaProA	525
Qy	1073	CATTC-----GCAGACTGGATCCGTAAGACGGCAGAGTCGCTCTATCGCATGTATTGT	1129
Db	525	laGluProLeuGluProValLysLysThrLysLeuValProValAspLeuGluValIleG	545
Qy	1130	CTCTCTCGAAAGGCTTAT-----	1149
Db	545	luserIleProValSerTyrAspValGlnLysPheHisnLeuLeuGlnMetGlnG	565
Qy	1150	-----AAGGAAGAGGACCAAGCCACCGTCAGATGACTTATTATTGACGACGAGCGC	1198
Db	565	luserAspAlaArgGluLysAlaLysAlaAspAlaLysAsnSer--LeuGluGluTyrV	584
Qy	1199	TCCTCGGTGTACGAGGTGGT-----CGTTTGCACAGTTTCCCAACGATGGCTA	1252
Db	584	alTyrGluMetArgAspLysValSerAspGlnTyrAlaGluPheIleThrProAlaAlaA	604
Qy	1253	CAATCTCGATGCTATCGCGGTATCCTCAAAATCGCTTCACCAAGTACAAAGACTACC	1312
Db	604	la-----AspGluPheArgSerValLeuThrSerThrGluAspIleTyrAspGluG	622
Qy	1313	TCCCTCGCTCGACCGTAAAGTGCTGCCGCCATGCTCGATATTCGTACGCCGCGTATCC	1372
Db	622	lyGluAspAlaGluArg-----AspValTyrGluLysArgLeu-	634
Qy	1373	CTGCCGACAGACTCCCGATATATTCAGAATGTATTCAGCAAGAANAATTCAAAGCGGACA	1432
Db	635	-----SerGluLeuLysAlaValIcLysThrProValValGluArgTyrArgGluSerGluT	653
Qy	1433	CGAAGAGTATGCAGACTTCTCTATTCGCAAGAGTCTGGCT-----CCTTATA	1480
Db	653	hrArgLysProAlaPheAspSerPheAspGlnSerIleMetArgValArgLysAlaTyrG	673
Qy	1481	CGCAC-----AAGTTCATGCCATGCTCAAGTCCATGGACAGGAAAAAGT	1525
Db	673	luAspTyrAlaAsnGlyGlyProThrTyrAlaHisLysSerLysGluMetGluLysV	693
Qy	1526	TTGCCAAGGCTATCGAG	1542
Db	693	allIleAsnAlaIleGlu	698
RESULT 4			
HK1	Yeast	STANDARD;	PRT; 1802 AA.
ID	HK1_YEAST		
AC	P41809;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	Hansenula MRK11 killer toxin-resistant protein 1 precursor.		
GN	HK1 OR YDR240W.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=YNN 295;		
RX	MEDLINE=94156857; PubMed=8113191;		
RA	Kasahara S., Yamada H., Mio T., Shiratori Y., Miyamoto C.,		
RA	Yabe T., Nakajima T., Ichishima E., Furuchi Y.;		
RT	"Cloning of the Saccharomyces cerevisiae gene whose overexpression		
RT	overcomes the effects of HM-1 killer toxin, which inhibits		
RT	beta-glucan synthesis.";		
RL	J. Bacteriol. 176:1488-1499(1994).		
CC	!- FUNCTION: COULD REGULATE BETA-GLUCAN SYNTHESIS. OVEREXPRESSION		
CC	PROVIDES RESISTANCE TO HM-1 KILLER TOXIN.		
CC	!- SUBCELLULAR LOCATION: Type I membrane protein (probable).		
CC	!- PTM: COULD BE O-GLYCOSYLATED IN SERINE/THREONINE RICH DOMAIN.		









QY 529 GTAGAGCCTTTCTATTCCACAAAGCAATACTTCTCATCTGCTACGATGATTTCAGGAC 588  
 Db 3879 MetTy-ProPheAsnProAsnSerAlaTy-  
 QY 589 GTTCGATGATTTGCTCTCCAGCTCTGTAGGTAAGTTCCGAGCGGATACGGAC--- 645  
 Db 3889 -----GluProAla-----PheGlnGlyAspGluAsnVal 3898  
 QY 646 -----AACTGGATG 654  
 Db 3899 ArgIleAspIleMetAspIleTyValLysGlyAsnLysIleTyTrpThrAsnTrp--- 3917  
 QY 655 TGGCCGCGTCACAGCGCGACTTCAGCGTATTCGCGGTGATCCGGT----- 702  
 Db 3918 -----HisThrGlyArgIleSerTyTyCysGluLeuProAlaSerAlaAlaSer 3934  
 QY 703 -----GCCACAAACCGG----- 714  
 Db 3935 ThrAlaSerAsnArgAsnArgGlnIleAspGlyValThrHisLeuAsnIleSer 3954  
 QY 714 ----- 714  
 Db 3955 GlyLeuLysMetProArgGlyIleAlaValAspTrpValAlaGlyAsnIleTyTrpThr 3974  
 QY 715 -----CCGCCGAATACAGCAAGGACATAAACCCCTATAAG 750  
 Db 3975 AspSerGlyArgAspValIleGluValAlaGlnMetLysGlyGluAsnArg----- 3991  
 QY 751 CCGGTTACTTCCGTCGCGTATCCAGCAAGGCTACAGGCTCAGCACTATGCGATGACC 810  
 Db 3992 -----LysThrLeuIleSerGlyMetIleAspGluProHisAlaIleVal 4006  
 QY 811 ATCGGTTTCCGGCGCAGTACGGATCGCTACCTCCTCTCTGTTGGGTGTGGAGATCGTATC 870  
 Db 4007 ValAsp---ProLeuArgGlyThrMetTyTrpSerAspTrpGly----- 4020  
 QY 871 GAAACAGAGAACATCTCGTATCGAAGTTCCGGGTATCAAGCAAGGCATCTCGAAGGAA 930  
 Db 4021 -----AsnHisProLysIleGluThrAlaAlaMet---AspGlyThrLeuArgGlu 4036  
 QY 931 GCCATGAGCCAGATCAGGTACCGGTATCAATAT-----GCCAGCAAGTAT 978  
 Db 4037 ThrLeuValGlnAsp-----AsnIleGlnTrpProThrGlyLeuAlaValAspTy 4053  
 QY 979 GCTCAGAGTCTAACTATTGGAAGAAATTCGATCGGTATGAACCGGCTCTCGTCTGCTT 1038  
 Db 4054 HisAsnGluArgLeuTyTrpAlaAsp-----AlaLysLeu 4065  
 QY 1039 GACGTATAGTTCGTAAAGCGTCGCGGAGAAAGAGACATTCGCACTGGATCCGTAAGAAC 1098  
 Db 4066 SerValIleGlySer-----IleArgLeuAsn 4074  
 QY 1099 GGC----- 1101  
 Db 4075 GlyThrAspProValValAlaIleAspAsnLysLysGlyLeuSerHisProPheSerIle 4094  
 QY 1102 -----AAGAGTGTGCTATGG-----GATGATTGTCTTCTCGAAAGGCT 1146  
 Db 4095 AspIlePheGluAspTyIleTyGlyValThrTyIleAsnAsnArgIlePheLysIle 4114  
 QY 1147 TATAAGGAAGGAGCCCAAGGCCCAACCGTGTAGATGACTTATTGTAGCGAGAGCGCTTCGGT 1206  
 Db 4115 HisLysPheGlyHisLysSer-----ValThrAsnLeuThrSerGlyLeuAsnHis 4131  
 QY 1207 GGTACCGAGGTGGTTCGTTTGGTTCGACATTTGGCCAAACCATTCGCTACAAATCCTGATGCT 1266  
 Db 4132 AlaThrAspValValLeuTyHisGlnTyLysGlnProGluValThrAsnPro----- 4149  
 QY 1267 CATGCCGTATCCTCAATCGCTTGACGACAAAGTACAAAGACTACCTC----- 1314  
 Db 4150 -----CysAspArgLysCysGluTrpLeuCysLeuLeuSer 4162

QY 1315 CCTCCTCGACCGCTAAGTGCTGCCCGCATCTCGATATTTGACCCGCGC----- 1366  
 Db 4163 ProSer-GlyProValCysThrCysProAsnGlyLysArgLeuAspAsnGlyThrCysVal 4182  
 QY 1367 ----GTATCCTCCGACACAGCTCCCGATATATTCAAGATGTAATCGACAGAATAATTC 1422  
 Db 4182 lLeuIleProSerProThrAlaSerAlaVal-ValProThrThrAspThrCysAspLeuV 4202  
 QY 1423 AAAGGCGACACGAAGAAGTATGCAGACTTCTGCTATTCGACAAAGAGTGTGGTTCCTTATAGC 1482  
 Db 4202 alCysLeuAsnGlyGlySerCysPheLeuAsnAlaArg-----L 4215  
 QY 1483 GACAAGTTCCATCCATGCTCAAGTCCATGCAGCAAGGAAAAGTTTGCAA----- 1532  
 Db 4215 ysGlnAlaLysCysArgCysGlnProArgTyArgGlnGlyGluArgCysGlnIleAsnGlnC 4235  
 QY 1533 -----GGCTATCAGAAAGATCCGCG-----AGTAGAGCTTTCCAGAGCGCTA 1575  
 Db 4235 ysSerAspTyTyCysGlnAsnGlyGlyLeuCysThrAlaSerProSerGlyMetProThrC 4255  
 QY 1576 ATAGCTCTCTCGCGCTATTTCAGG---CGATCGGATGGCAA---TGC-----CTAT 1623  
 Db 4255 ysArgCysProThrGlyPheThrGlySerArgCysAspGlnGlnValCysThrAsnTyrc 4275  
 QY 1624 GCCATTGAGAAAGGCAAGCGTCTTT-----CTTTGCCGTTTCGCT 1665  
 Db 4275 ysHisAsnAsnGlySerCysThrValAsnGlnGlyAsnGlnProAsnCysArg----- 4292  
 QY 1666 GAGATGTACCCCGGCGTCTCTCCGAGGATGCCAACTTCACCATCGGTATGAGGTAC 1725  
 Db 4293 -----CysProProThrPheIleGlyAspArgCysGlnTyGlnGlnCysPheAsnTyrc 4311  
 QY 1726 GGCTCCATCAAGGATATGAACCGCAGCGTCCCTGGTACAACTATCATACGAC 1781  
 Db 4311 ysGluAsnAsnGlyVal-----CysGlnMetSerArgasp 4322

RESULT 6  
 ID YM96\_YEAST STANDARD; PRT; 1140 AA.  
 AC Q04893;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.  
 GN YMR317W OR YMR924.09.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; 254141; CAA90835.1; -  
 DR SGD; S0004936; YMR317W.  
 KW Hypothetical protein; Repeat.  
 SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Alignment Scores: 0.195 Length: 1140  
 Pred. No.: 137.50 Matches: 131  
 Score: 36.73%  
 Percent Similarity: 36.73%

Best Local Similarity: 21.20% Mismatches: 246  
Query Match: 3.60% Indels: 145  
DB: 1 Gaps: 21

US-10-008-355-1 (1-2139) x YM96\_YEAST (1-1140)

```
Qy 533 AGCCTTTCTATTCACACGAATCTCTCATCGTCTACGATGTATTCAGGACGTC 592
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 SerThrSerProAlaSerThrIleSerGluThrLeuProPheSerSerThrIle 252
Qy 593 GTATGGTATTTCCTCCAGCTCTGTAGTAAAGTTCGGAGGGATACGACAACTGGA 652
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 LeuSerIleThrSerSerPro-----ValSerSerGluAlaProSerAlaThrSer 269
Qy 653 TGTGGCCGCTCACACGGCGCACTTCAGCGTATTCCGGGTGTATG-----CCG 700
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 SerSerValSerSerGluAlaSerSerThrSerSerValSerSerGluAlaPro 289
Qy 701 GTGGCGCAACCGGC-----CGGCCGAATACAGCAAGGACAATA----- 739
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 290 LeuAlaThrSerSerValSerSerGluAlaProSerSerThrSerSerValSer 309
Qy 740 -----ACCTATTAAGCCGTTTACTTCGCTGCGGTATCCATCGAAGGCTACAGGCTG 793
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 SerGluAlaProSerSerThrSerSerVal-----SerSerGluIle 324
Qy 794 ACGACTATGCCATGACCATCGGTTTCCGGGCGAGTACGGATCGCTACCTCTT--- 850
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 SerSerThrThrSerSerValSerSerGluAlaProLeuAlaThrSerSerVal 344
Qy 851 -----GGGTGTGGAAAGATCGTATCGAATAACGAGA 880
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345 SerSerGluAlaProSerSerThrSerSerValSerSerGluIleSerSerThr 364
Qy 881 ACAATCCTCGTATCAAG-----TTCGGCGTATCAACGAGGCATCTGGAGGAG 931
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 SerSerSerValSerSerGluAlaProLeuAlaThrSerSerValSerSerGluAla 384
Qy 932 CCATGAGCGCAGATCAGGCTACCGGTATCAATATGCCAGCAAGTATGCTCAGAGTGCTA 991
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 ProSerSerThrSerSerValSerSerGluAlaProSerSerThrSerSerSerVal 404
Qy 992 ACTATTGGAAGAAATTCGATCGGTATGAACCGGGTCTCGTCTGTACGTAGTAGTGC 1051
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 405 SerSerGluAlaProSerSerThrSerSerValSerSerGluIleSerSerThrLys 424
Qy 1052 GTAAGCGTCCGAGGAAAGACATTCGACAGCTGGATCCGTAAGAACGGCAGAGTGTG 1111
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 SerSerValMetSerSerGluValSerSerAlaThrSerSerLeuValSerSerGluAla 444
Qy 1112 TCTATGGCGATGTATTGTCTCTCGAAAAGGCTTATAAGGAAGGACCAAGGCCAAC 1171
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 ProSerAlaIleSerSerLeuAlaSerSerArgLeuPheSerSerLysAsnThrSerVal 464
Qy 1172 GTGAGATGACTTTATTTAGCAGAGGCTCTTCGGTGGTACCGAGGTGTTGGTTTGCAC 1231
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 465 ThrSerThrLeuValAlaThrGluAlaSerSerValThrSer----- 478
Qy 1232 AGTTTGCCAACGCATGGCTACAAATCTGTATGCTATGCGC-----GTATCC 1279
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 -----SerLeuArgProSerSerGluThrLeuAla 488
Qy 1280 TCAATCCTTTGACGACAGATCAAAAGACTACCTCCCTCGCTCGACCGTAGGTGTGTC 1339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 489 SerAsnSerIleIleGluSerSerLeuSerThrGlyThrAsnSerThrVal----- 505
Qy 1340 CCGCATGCTCGATTTGTAGCGGGGTATCCCTGCGGCAAGACTCCCGCATATATCA 1399
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 -----SerThrThrSerAlaAlaSerSerThrLeuGlySer 518
Qy 1400 AGAATGTAATCGACAAGAAATTCAAAGCGGACAGCAAGACTATCCAGACTTCG----- 1453
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 519 LysValSerSerSerAsnSerArgMetAlaThrSerLysThrSerSerSerSerAsp 538
```

```
Qy 1454 -----TATTCGACAAGAGTGTGGTTC 1474
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 LeuSerLysSerSerValIlePheGlyAsnSerSerThrValThrSerProSerAla 558
Qy 1475 CTTATAGCGACAAGTTCATGCCCATGTCAAGTCCATGG-----ACAAGGAAAGT 1525
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 559 SerIleSerLeuThrAlaSerProLeuProSerValTrpSerAspIleThrSerSerGlu 578
Qy 1526 TTGCCAAGGCTATCAGAAAAGATCCGCGCAGTAGAGCTTTCCAAAGAGCTATAGCTGCTG 1585
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 579 AlaSerSerIleSerSerAsnLeuAlaSerSerSerAlaProSerAsp----- 594
Qy 1586 CTCGCGCTATTACGCGCGATCGCATGGCAATGCCTATGCCTAGCAAGGCGCAAGGCTC 1645
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 -----AsnAsnSerThrIle 599
Qy 1646 TTTTCTTTGCGGTTTTCGTTGAGATGATACCCCGGACGTCCTCCGAGCGATCCCAACT 1705
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 AlaSerAlaSerLeuIleValThrLysThrLysAsnSerValSerSerIleValSer 619
Qy 1706 TCACCATGCGTATGAGTACGCTCCATCAAGGGATATGAACCGCAGGACGCTGCTGCT 1765
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 620 Ser-----IleThrSerSerGluThrThrAsnGluSerAsnLeuAlaThr---SerSer 636
Qy 1766 ACAACTATCATACGACAG-----GCAAGGCGCTATTGGAGAAGCAGGATCCTAAGA 1816
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 ThrSerLeuLeuSerAsnLysAlaThrAlaArgSerLeuSerThrSerAsnAlaThrSer 656
Qy 1817 GCGATGAGTTTCCGCTACAGGAGATATCTCCGACCTCT----- 1855
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 657 AlaSerAsnValProThrGlyThrPheSerSerMetSerSerHisThrSerValIleThr 676
Qy 1856 -----TCCGCAACCAAAACTATGTCGCTATCCGAGAAGCGTCAGCTCCATATCG 1906
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 677 ProGlyPheSerThrSerSerAlaSerLeuAlaIleAsnSerThrValValSerSerSer 696
Qy 1907 CTT-----TCCTATCGAACA-----ACGACATCACGGCGGTAACT 1942
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 LeuAlaGlyTyrSerPheSerThrProGluSerSerProThrThrSerThrLeuValThr 716
Qy 1943 CCGTATGCGCGTATTGATGAAGAACGGCGCTCTGATCGGTC----- 1984
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 SerGluAlaProSerThrValSerSerMetThrThrSerAlaProPheIleAsnAsnSer 736
Qy 1985 -----TTGCTTTTCGATGGCAACTGGGAAGCTA 2011
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 737 ThrSerAlaArgProSerProSerThrAlaSerPheIleThrGluSerThrSerSerIle 756
Qy 2012 TGAGTGTGATCGATCGAGTTCGAACCCGATCTCGACGCGACCAATCAGGCTGGACATCGCT 2071
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 SerSerValProLeuAlaSerSerGlyAspValThrSerSerLeuAlaAlaHisAsnLeuThr 776
Qy 2072 ACGTTCCTTCATGATTCACAAATGGGTGTCAGTCCCGCTCTCATCCAGAGC 2125
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 ThrPheSerAlaProSerThrSer-----SerAlaGlnLeuValSerLysSer 792
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7

PT1\_CHLTR

ID PT1\_CHLTR STANDARD; PRT; 571 AA.

AC 084340;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9)

DE (Phosphotransferase system, enzyme 1).

GN PPS1 OR CT336.

OS Chlamydia trachomatis.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=813;

QX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D/UW-3/CX;

RX MEDLINE=99008009; PubMed=9784136;  
 RA Stephens R.S., Kalman S., Lammell C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis";  
 RL Science 282:754-759(1998).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC -TRANSPORT SYSTEM. ENZYME 1 TRANSFERS THE PHOSPHORYL GROUP FROM  
 CC PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).  
 CC ENZYME 1 IS COMMON TO ALL PTS (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + protein L-histidine -  
 CC pyruvate + protein N(pi)-phospho-L-histidine.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AE001306; AAC67931.1; -  
 DR InterPro; IPR000121; PEP-utilizers.  
 DR Pfam; PF00391; PEP-utilizers; 1.  
 DR Pfam; PF02896; PEP-utilizers; C; 1.  
 DR PRODOM; PD000940; PEP-utilizers; 1.  
 DR PROSITE; PS00370; PEP-ENZYMES\_PHOS\_SITE; FALSE\_NEG.  
 DR PROSITE; PS00742; PEP-ENZYMES\_2; FALSE\_NEG.  
 KW Phosphotransferase system; Transferrase; Kinase; Sugar transport;  
 KW Phosphorylation; Complete proteome.  
 FT MOD\_RES 207 207 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 571 AA; 63786 MW; 73EBFE7C457067 CRC64;

## Alignment Scores:

Pred. No.: 0.362 Length: 571  
 Score: 132.00 Matches: 115  
 Percent Similarity: 35.37% Conservative: 88  
 Best Local Similarity: 20.03% Mismatches: 186  
 Query Match: 3.33% Indels: 185  
 DB: 1 Gaps: 24

US-10-008-355-1 (1-2139) x PT1\_CHLTR (1-571)

QY 1635 CTTCTCAATGCGATAGGATGGCCATCGCATCGCGCTGANTAGCGGCGAGCAGCTAT 1576  
 Db 22 ILeValProGlyIleGlyLeuGlyLysAlaLeu----- 32  
 QY 1575 TACGCTCTTGGAAAGCTCTACTCGCGGATCTTTCTCGATAGCCTTGGCAAACTTTTCCTT 1516  
 Db 33 --LeuLeuGlyLysSerSerLeuArgIleArgGluLeuThrLeuProGlnGluVal 51  
 QY 1515 GTCCATGACTTGTAGCAT-----GGCATGGAACCTTGTCTCGT 1480  
 Db 52 GluHisGluIleSerArgTyTyTyLysAlaLeuLysArgSerArgSerAlaAla 71  
 QY 1479 ATAAGGACACACTCTGTGCGAA-----TACGAAGTCTGCATCTTCTTCGTGTC 1429  
 Db 72 LeuGluLysGluAlaLysGlyLysGlnGlyTyrglnGluIleAlaSerIleLeuGlnAla 91  
 QY 1428 GCCTTTGAATTTCTGTGTACATCTT---GAATATATCGGGAGCTTGTGCGGAGG 1372  
 Db 92 HisLeuGluIleIleLysAspProLeuLeuThrGluGluValValLysThrIleArgLys 111  
 QY 1371 GATACGGCGGCGTACAATATACGACATGGCGGCGAGCACCTTACGCTGCGAGGGGAG 1312  
 Db 112 AspArg-----LysAsnAlaGlu 117  
 QY 1311 GTAGCTTTGTACTTGTGCTACGCGATTTGAGGATACCGCGCATGAGCATCAGG-----1258  
 Db 425 SerGluAsnIleLeuTrpGlySerMetIleGluIleProSerAlaValTrpMetIleGlu 444

Db 118 PheValPheSerSerValMetGlyGluIleGlySerLeuGlySerLeuGlyAlaValGlnLysThr 137  
 QY 1257 -----ATTGTAGCAATAGCTTTCGC 1237  
 Db 138 ThrAlaThrArgValAspArgValGlnAspIleHisAspIleSerAsnArgValIleGly 157  
 QY 1236 AAACGTGTC-----AAACGAACACCTCGGTACCCAGAGCGTCTC----- 1192  
 Db 158 HisLeuGlySerGlnHisLysSerSerLeuGlyGluPheAspGlnAsnLeuIleValPhe 177  
 QY 1192 ----- 1192  
 Db 178 SerGluGluLeuThrProSerGluAlaAlaAsnAlaAsnProGluTyIleArgGlyPhe 197  
 QY 1191 -----GCTCAAAATAAGTCTCATCGGTTGGCTTGGCTTCCTTCCTTATA 1147  
 Db 198 ValSerLeuGluGlyAlaLysThrSerHisThrAlaIle-----Val 211  
 QY 1146 AGCCTTTTCGAGAGAACAAATACATCGCC----- 1117  
 Db 212 SerLeuAlaLysAsnIleProTyValAlaAlaAsnPheThrThrGluLeuTrpAspThrIle 231  
 QY 1116 -----ATAGACACACTCTTGGCGTCTTACGGATCCAGTCTCGCAATGCTCTTTC----- 1066  
 Db 232 LysGluPheSerGlyThrLeuValLeuIleAsnGlyAspLysGlyGluIleThrPheAsn 251  
 QY 1065 -----CTCGGCGAGCTTACGACCTATCACGCTCAAG-----ACGAGC 1030  
 Db 252 ProGlnLeuSerThrIleGlnThrTyTyTyArgLysGlnAlaSerValSerValThrVal 271  
 QY 1029 GAGACCGCGGTTTCATACCGATCTTCCAAATAGTAGTACTCTGAGCATCTGCTGCT 970  
 Db 272 ProValGlnValGlnThrGlyLysAsnLeuProLeuIleSer-LeuSerAla----- 288  
 QY 969 GGCATATTTGATACGGTAGCGCTGATCTGCGCTCATGCTCTCCAGATGCTCTGCTT 910  
 Db 289 -----GlnIleValSerThrGluGluLeuProMetIle 299  
 QY 909 G-----ATACCGGAACTTCGATACGAGGATTTGTTCTCTTTCGATACGATCTTC 859  
 Db 299 eGluArgGluSerProGlyThrSerVal---GlyLeuPhe----- 311  
 QY 858 CACACCCCAAGAAGTGAGGTAGCGATCGTACTGCCCGGGAACCGATGGTTCATGGCATA 799  
 Db 312 -----ArgSerGluPheMetAlaPhe 318  
 QY 798 GTCGTACGCTTGTAGCCTTGCATGCGATACGCGACCGCAAGTAACCGGCTTATAGGTTT 739  
 Db 318 eSerLeuGlyArgLeuProCysValGlu-GluGlnAlaAspGlnTyAlaGlnLeuValG 338  
 QY 738 AT-----T 736  
 Db 338 InPheGlnCysSerAspIleHisValLeuArgLeuPheAspPheGlyLysAspLysGluC 358  
 QY 735 GTCCTTGTGTTATTCGCGCGCGCGGTTGTCGGCACCGCGCATACAGCGGAATACGCTGAA 676  
 Db 358 ysProCys-----IleSerSerHisArgSerValArgTrpLeu-LeuGlu 373  
 QY 675 GTCGCCCGTGTGACGCGGCCACAT---CCAGTTGTCGTATCGCTCCGCAACTACCTAC 619  
 Db 374 GlnGluLysValLeuLysGluGlnLeuGlnAlaIleAlaIleValSer----- 389  
 QY 618 ACAGCTGGGAGGAGCAATACCATACGAACTGCTTGAATACATCATGATACGATGAGGAA 559  
 Db 390 ArgIleGlyArgLeuLysVal-----LeuIleProGlyValIleAspAlaSerGlu 406  
 QY 558 GTATTCGTTGTTGAATAGAAAGCTCTACGATCGAGAGTTGGTTCTGCTGCAATTTC 499  
 Db 407 IleAlaLeuVal-----LysArgLeuPheGlnGluGluIleArgLeuLysGlyIle 424  
 QY 498 TTTTTCGCCAGTCTTCTGGCATACCTCTGAGCTTGGCGAGACGCTCCATCTCTCAGT 439  
 Db 425 SerGluAsnIleLeuTrpGlySerMetIleGluIleProSerAlaValTrpMetIleGlu 444



Db 386 IleSerThrThrIleProSerPheSerMetSerThrTyrPheThrThrValSerGlyVal 405  
Qy 1138 GAAAGGCTTATAGGAGGAGGCAAGCCACCTGAG-----ATGACT 1182  
Db 406 ThrThrMetThrThrThrTpCysProTyrSerSerGluSerGluThrSerThrLeuThr 425  
Qy 1183 TATTGAGCGAGCGCTTCGGTGGTACCGAGGTGTTGCTTTGACAGTTGCGCAAC 1242  
Db 426 SerMetHisGluThrValThrThrAspAlaThrValCysThrHisGluSerCysMetPro 445  
Qy 1243 GCATTGCTCAAACTCTGATGCTC-----ATCCCGGTATCTCAAACTGCTGAGCAG 1296  
Db 446 SerGlnThrThrSerLeuThrThrSerSerIleLeuMetSerThrLysAsnValAlaTh 465  
Qy 1297 AAGTCAAAAGACTACCTCCCTCCCTCGACCGTAAGTGC-----TGCCGCGCATGCTC 1350  
Db 465 rSerValSerThrSerThrValGluSerSerTyrAlaCysSerThrCysAlaGluThrSe 485  
Qy 1351 GATATTGTACCGCGGTATCCCTGCGCACAGCTCCCGATATATTCACAAATGTAATC 1410  
Db 485 rHisSerTyrSerSerValGlnThrAlaSerSerSerValThrGlnGlnThrThrSe 505  
Qy 1411 GACAGA-----AATCAAGGCGAC 1431  
Db 505 rThrLysSerTrpValSerSerMetThrThrSerAspGluAspPheAsnLysHisAlaTh 525  
Qy 1432 ACGAAGAGTATGACGACTTCGTATTCGACAAAGAGTGTGTTCTTATATGCGACAAGTTC 1491  
Db 525 rGlyLysThrHisValThrSerSerGlyThr--SerThrIleSerThrSerValSerGlu 544  
Qy 1492 CATGCCATGCTCAAGTCCATGAGGCAAGAAAGTTTGCAGAGGTATCGCAAGATCCG 1551  
Db 545 AlaThrSerThrSerSerIleAspSerGluSer-----GlnGluGlnSer 559  
Qy 1552 GCATGAGGCTTCCAGAGCGTAACTACCTGCTCGCTGCTATTCAGCGCGATGCCATG 1611  
Db 560 SerHisLeuLeuSerThrSerValLeuSerSerSer---SerLeuSerAlaThrLeuSer 578  
Qy 1612 GCCAATGCTATGCTCATGAGAGGCAAGGCTCTTTCTTTGCGGTT-----1660  
Db 579 SerAspSerThrIle-LeuLeuPheSerSerValSerSerValGluGlnSerPr 598  
Qy 1661 -----TGCCTGAGATGATCCCGGACGCTCTGCGGAGCGATGCCAACTTC 1707  
Db 598 oValThrThrLeuGlnIleSerSerThrSerGluIleLeu-----GlnProThrSe 615  
Qy 1708 ACCATGCTATGACTACGGCTCCATCAAGGATATGACCGCAGGCGTGGCTGTAC 1767  
Db 615 r-----Se 616  
Qy 1768 AACTATCATGACGAGGCAAGGCGGTATTGGAGAGCAGGATCCTAAGAGCGATGTTT 1827  
Db 616 rThrAlaIleAlaThrIleSerAlaSerThrSerSerLeuSerAlaThrSerIleSerTh 636  
Qy 1828 CCGTACAGGAGATATCTCGACCTCTCCGACCAAAAACTATGCTGCTATGCCGAG 1887  
Db 636 rProSerThrSerValGluSerThrIleGluSerSer-----SerLeuThrPr 652  
Qy 1888 AACGTCAGCTCCATATCGCTTCTCTATCGAACACGATCAGCGGCGGTAACTCCGCT 1947  
Db 652 oThrValSerSerIlePheLeuSer-----SerSerSe 663  
Qy 1948 AGCCCGTATTGATAAGACGCGCTGTATGCTGCTTCTGTTTCGATGCGCAACTGGAA 2007  
Db 663 rAlaProSerSerLeuGlnThrSerValThrThrThrGluValSerThrThrSerIleSe 683  
Qy 2008 GCTAGTGTGTCATCGAGT 2029  
Db 683 rIleGlnTyrGlnThrSerSer 690

RESULT 9

YS89\_CAEEL

ID YS89\_CAEEL

STANDARD; PRN; 3178 AA.

AC Q09624; Q09625; Q0969D4;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein ZK945.9 in chromosome II.  
GN ZK945.9/ZK945.10.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Wilkenson-Sproat J.;  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Durbin R.;  
RC Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z48544; CAB70192.1; JOINED.  
DR EMBL; Z48582; CAB70192.1; JOINED.  
DR EMBL; Z48582; CAB70201.1; JOINED.  
DR EMBL; Z48544; CAB70201.1; JOINED.  
DR WormPep; ZK945.9; CE25697.  
DR InterPro; IPR002111; Cat\_channel\_TrpL.  
DR InterPro; IPR001024; Lipoxigenase\_LH2.  
DR InterPro; IPR000636; M-channel\_nlg.  
DR InterPro; IPR000203; PKD\_cys\_rich.  
DR Pfam; PF01825; GPS; 1.  
DR Pfam; PF01477; PLAT; 1.  
DR Pfam; PF00520; Ion\_trans; 1.  
DR SMART; SM00303; GPS; 1.  
DR SMART; SM00308; LH2; 1.  
KW Hypothetical protein; Transmembrane.  
FT DOMAIN 266 1196 SER/THR-RICH.  
FT DOMAIN 1105 1241 GLY/SER-RICH.  
FT TRANSMEM 13 30 GPS.  
FT TRANSMEM 51 73 POTENTIAL.  
FT TRANSMEM 2139 2161 POTENTIAL.  
FT TRANSMEM 2348 2367 POTENTIAL.  
FT TRANSMEM 2390 2412 POTENTIAL.  
FT TRANSMEM 2451 2468 POTENTIAL.  
FT TRANSMEM 2483 2505 POTENTIAL.  
FT TRANSMEM 2567 2589 POTENTIAL.  
FT TRANSMEM 2836 2858 POTENTIAL.  
FT TRANSMEM 2939 2961 POTENTIAL.  
FT TRANSMEM 2976 2998 POTENTIAL.  
FT TRANSMEM 3038 3060 POTENTIAL.  
SQ SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;

Alignment Scores:

Pred. No.: 147 Length: 3178  
Score: 126.50 Matches: 148  
Percent Similarity: 31.05% Conservative: 70  
Best Local Similarity: 21.08% Mismatches: 227  
Query Match: 3.31% Indels: 258  
DB: 1 Gaps: 25

US-10-008-355-1 (1-2139) x YS89\_CAEEL (1-3178)

Qy 346 GCTGAGGAGCTTCGATTCGCGGT-----CTTCTCGTG 378

```
Db 186 GlyAsnGluMetAlaLeuLeuGlyTyrArgGluLysCysGluSerGlyGluIleAsnGlu 205
Qy 379 AAGTATCTCGCAAGCATCGTGAAG-----GTAACG 408
Db 206 GluTyrAlaArgMetCysLysArgProTyrArgSerGluLysSerThrAlaIleSer 225
Qy 409 GACAAG-----GTAGAAGGACAG---CTCAAGGGTATCATCTACGAGAGATGGAG 453
Db 226 AspSerGlnGlyValTyrTyrAspGlyGlnValLeuLysGlyValArgAlaLysGlnPhe 245
Qy 454 CGTCTGCCCAAGCTCAGAGGATGATGCCAAGAACTGGCCAAAGAAAGAAATGCGACGAG 513
Db 246 SerMetArgThrSerGlySerProThrLeuArgArgMetLysArgAspAlaGlyAsp-As 265
Qy 514 AACCC-----AACTCTGCATCTAGAGCCCTTCTATTATCCAAACCAACCAATAC 558
Db 265 nThrCysAspTyrThrIleGluSerThrSerThrSerThrThrThrThrThrThr 285
Qy 559 TTCCTCATCGTCTACGATGATTTCAAGGACGTTCTGATGGTATTTGCTCTCCCTCCAGCTCT 618
Db 285 rValThrSerThr----- 289
Qy 619 GTAGTAAGTTCGGAGGCGATACGGACAACCTGGATGTCGGCCCGTCACACGGGCGACTTC 678
Db 290 ---ValThrSerThrThrThrThrValProThrSerThrSerThrValThrThrAlaMetSe 308
Qy 679 AGCGTATTCGGCGTGTATGCGCGGTGCCGACACACCGCGCGCGGCGGATACAGCAAGCAAT 738
Db 308 rThr-----SerThrSerThrProSerThrSerThrThrThrThr 320
Qy 739 AAACCCCTATAAGC---CGGTTTACTTCGCTGCGGTATCCATGCAAGGCTACAAGGCTGAC 795
Db 320 eLusSerThrSerThrPheThrSer----- 329
Qy 796 GACTATGCCATGACCATCGGTTTCCCGGCGAGTAGCGATCGCTACCTCACCTCTTGGGGT 855
Db 330 -----ThrAlaSerThrSer----- 334
Qy 856 GTGGAAGATCGTATCGAAGAACGAGAACAACTCTCGTATCGAAGTTCGGCGTATCAAGCAA 915
Db 335 -----ThrSerSerThrSerThrThrGlnSerSerSerThrThrThrSerSe 351
Qy 916 GCATCTCGAAGGAGGAGCCATGAGCGCAGATCAGGCTACCGTATCAAAATATGCCA----- 970
Db 351 rProSerSerThrThr-----LeuSerThrSerIleProThrThrThrThrProGluI 369
Qy 971 -GCAAGTATGTCACAGTGTAACTATTGGGAAGAAATTCGATCGGTA-----TGAACCGC 1023
Db 369 eThrSerThrLeuSerSerLeuProAspAsnAlaIleCysSerTyrLeuAspGluThrTh 389
Qy 1024 GGTCTCGCTCTGCTTGACGTGATAGTTCGTAAGCGTGCAGGAGGAGGAGGATTCGCAGAC 1083
Db 389 rThrSerThrThrPheThrThrThrMetLeuThrSerThrThrThrGluGluProSerTh 409
Qy 1084 TGGATCCGTAAGACGGCAAGAGTGCTGCTATGCGCGATGATTGCTCTCTCTCGAAAG 1143
Db 409 r---SerThrThrThrThrGluValThrSer-----ThrSerSerTh 422
Qy 1144 GCTTATAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1203
Db 422 rValThrThrThrGluProThrThrThrLeu----- 432
Qy 1204 GGTGTACCGAGGTGCTCGTTTTCGACAGTTTGCAACGCGATTGGCTACAAATCCTGAT 1263
Db 432 ----- 432
Qy 1264 GCTCATCGGCTATCTCAAAATCGCTTGACGACAGTACAAAGACTACCTCCCTCT---CG 1320
Db 433 -----ThrThrSerThrAlaSerThrSerThrThrGluProSerThr 446
Qy 1321 CTCGACCGTAAGGTCTCGCCCGCATGCTCGATATTGACCGCGCGGTATCCCTCGCGAC 1380
|||||
```

```
Db 446 rSerThrVal-----ThrThrSerProSe 454
Qy 1381 AAGCTCCCGCATATATCAAGAATGTATTCACAGAATTCACAGGCGACACGAAGAAG 1440
Db 454 rThrSerProValThrSerThrValThrSerSerSerSerSerThrThr----- 471
Qy 1441 TATCAGACTTCGATTTCGACAGAGGTGTGTTCTTATACGACAAAGTTCCATGCCAT 1500
Db 472 -ValThrThrProThrSerThrGlu-----SerThrSerThrProSe 486
Qy 1501 CTCAGTCCATGGACAGAAAGTTTGCACAAAGGCTATCGAG----- 1542
Db 486 rSerThrValThrThrSerThrThrAlaProSerThrSerThrThrThrGlyProSerSe 506
Qy 1543 ---AAAGATCCGCGAGTAGAGCTTCCAAAGCGGTATAGCTGCTCGCTCGCTATTCA 1598
Db 506 rSerSerThrProSerSerThrAlaSerSerSerValSerSerThrAlaSerSerThrGl 526
Qy 1599 GCCGATCGCGATGGCC----- 1614
Db 526 nSerSerThrSerThrGlnSerSerThrThrThrLysSerGluThrThrThrSerSe 546
Qy 1615 -----AATGCTATGCCATTGACAGGCGGCAAGCGTCTTTTCTTGGCGG 1658
Db 546 rAspGlyThrAsnProAspPheTyrPheValGluLysAlaThrThrThrPheTyrAspSe 566
Qy 1659 TTTGCGTGAGATGTACCCCGGACGCTCTCCGCGAGCGATGCCAACTTCACCATCGGTAT 1718
Db 566 r-----ThrSerValAsnLeuThrLeuAsnSe 575
Qy 1719 GAGCTACGGCTCCATCAAGGATATGAA----- 1746
Db 575 rGlyLeuGlyIlele---GlyTyrGlnThrSerIleGluCysThrSerProThrSerSe 594
Qy 1747 -----CCGAGGAGGCTGCTCGGTACAC----- 1770
Db 594 rAsnTyrValSerThrThrLysAspGlyAlaCysPheThrLysSerValSerMetProAr 614
Qy 1770 ----- 1770
Db 614 gLeuGlyGlyThrTyrProAlaSerThrPheValGlyProGlyAsnTyrThrPheArgAl 634
Qy 1771 -----TATCATACGACAGGCAAGGCGGTAT----- 1795
Db 634 aThrMetThrThrAspAspLysLysValTyrTyrThr-TyrAlaAsnValTyrIleGlnG 654
Qy 1796 --TGAGAGGAGGATCCTAAGAGCGATGAGTTGCCGTACAGGAGATATCTCGACCT 1853
Db 654 lufyrSerSerThrThrIleGluSerGluSerSerThrSerAlaValAlaSerSerThrS 674
Qy 1854 CTCCCGCACAAACTATGCTGCTATGCTCGGAGACG-----GTCAGCT 1898
Db 674 erSerThrProSerThrProSerSerThrLeuSerThrSerThrValThrGluProSerS 694
Qy 1899 CCATATCGCTTTCATCGAACAACGACATCAGCGGCGGTAACTCCGCTAGCCCGCTATT 1958
Db 694 erThrArgSerSerAspSerThrThrThrSerAlaGlySerThrThrThrLeuGlnGluS 714
Qy 1959 CGATAAGAACCGCGCTGATGCTGCTGCTTTCGATGGCAACTGGGAGGATGAGTGG 2018
Db 714 erThrThrThrThrGluSerThrThrThrThrThrThrThrThrThrThrThrThrThr 734
Qy 2019 TGACATCGAGTTCGAACCCGATCTCGACGCAATCAGCGTGCAGCATCCGCTACGTTCT 2078
Db 734 erThrThrSerSerSerPro-----SerSerThrThrAlaAspSerThrThrLeuS 752
Qy 2079 CT 2080
Db 752 er 752
RESULT 10
COLA_CLOPE
ID COLA_CLOPE STANDARD; PRT; 1104 AA.
```



AC P43153;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Microbial collagenase precursor (EC 3.4.24.3) (120 kDa collagenase).  
 GN COLA OR CPE0173  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 87-113.  
 RC STRAIN=NCIB 10662;  
 RX MEDLINE=94110220; PubMed=8282691;  
 RA Matsushita O., Yoshihara K., Katayama S.-I., Minami J., Okabe A.;  
 RT "Purification and characterization of Clostridium perfringens 120-  
 RT kilodalton collagenase and nucleotide sequence of the corresponding  
 RT gene";  
 RL J. Bacteriol. 176:149-156(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / Type A;  
 RX PubMed=11792842;  
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 RT flesh-eater";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 RN [3]  
 RP SEQUENCE OF 1073-1104 FROM N.A.  
 RC STRAIN=NCIB 10662;  
 RA Matsushita O.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Digestion of native collagen in the triple  
 CC helical region at Xaa-I-Gly bonds. With synthetic peptides, a  
 CC preference is shown for Gly at P3 and P1; Pro and Ala at P2 and  
 CC P2; and hydroxyproline, Ala or Arg at P3.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M9.  
 CC -1- SIMILARITY: CONTAINS 1 PKD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D13791; BAA02941.1; -;  
 DR EMBL; AP003185; BAB79879.1; -;  
 DR EMBL; D50309; BAA08848.1; -;  
 DR MEROPS; M09.002; -;  
 DR InterPro; IPR002169; Micollptase.  
 DR InterPro; IPR000601; PKD\_domain.  
 DR InterPro; IPR000130; 2p\_MTPeptidse.  
 DR Pfam; PF01752; Peptidase\_M9; 1.  
 DR Pfam; PF01752; Peptidase\_M9; 1.  
 DR PRINTS; PR00931; MICOLLPTASE.  
 DR SMART; SM00089; PKD; 1.  
 DR PROSITE; PS00093; PKD; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Zinc; Zymogen; Signal; Complete proteome.  
 FT SIGNAL 1 39  
 FT PROPEP 40 86  
 FT CHAIN 87 1104  
 FT DOMAIN 774 862  
 FT METAL 502 503  
 FT ACT\_SITE 503 503  
 FT METAL 506 506  
 FT METAL 38 38  
 FT CONFLICT 722 722  
 FT CONFLICT 748 748  
 FT CONFLICT 945 945

FT CONFLICT 970 970 T -> A (IN REF. 1).  
 FT CONFLICT 987 987 A -> E (IN REF. 1).  
 FT CONFLICT 1098 1098 I -> T (IN REF. 1 AND 3).  
 SQ SEQUENCE 1104 AA; 125935 MW; F4B7377194ED021C CRC64;  
 Alignment Scores:  
 Pred. No.: 2.08 Length: 1104  
 Score: 122.00 Matches: 141  
 Percent Similarity: 31.46% Conservative: 105  
 Best Local Similarity: 18.03% Mismatches: 278  
 Query Match: 3.19% Indels: 258  
 DB: 1 Gaps: 36  
 US-10-008-355-1 (1-2139) x COLA\_CLOPE (1-1104)  
 QY 85 TGGCTCTCAACGAACCAATCAAGGAGAACTCTG-----GATCGAATCGGT 129  
 DB 419 TrpLaserLysGluValLysAlaGlnPheMetArgValValGlnAsnAspLysAlaLeu 438  
 QY 130 GAGCTCGGCTTTACGCTCCGCTGGATTGCTCTACAGTTTCGACAAAGCCGTCATTGCC 189  
 DB 439 GluGluGlyAsnProAspAspIleLeuThrValValIleTyrAsnSerProGluGluTyr 458  
 QY 190 AATCCCGTGGTTATCTTCGGTGGCGGATGTACC-----GGTATCACAGTGTCCGAT 240  
 DB 459 LysLeuAsnArgIleIleAsnGlyPheSerThrAspAsnGlyGlyIleTyrIleGluAsn 478  
 QY 241 CAGGCGCTGATCTTTACCAACACCACCACCTCGGATACGGTGCTATCCAGAGC---CAAAGC 297  
 DB 479 IleGlyThrPhePheThrTyrGluArgThrProGluGluSerIleTyrLeuGluGlu 498  
 QY 298 ACGGTGCATCACGAC-----TATCTGCGCGATGGTTTCTTCGACGATGGGT 348  
 DB 499 LeuPheArgHisGluPheThrHisTyrLeuGlnGlyArgTyrVal-----513  
 QY 349 GAGGAGCTTCGATTCGCGGTCTT-----TCCGTGAAGTATCTCGCAAGATCOTG 399  
 DB 514 -----ValProGlyMetTrpGlyGlnGlyGluPheTyrGlnGluGlyValLeu 529  
 QY 400 AAGTAAGGAGGAGTAGAAGGACAGCTCAAGGATATCACTGACGAGATGAGCGCTCG 459  
 DB 530 ThrTrpTyrGluGluGlyThrAlaGluPheAlaGlySerThrArgThrAspGlyIle 549  
 QY 460 CGAAAGCTCAGGAGGTATGCCAAGACTGCGCAAAAGAAAATGACAGCAGCAACAA 519  
 DB 550 LysProArgLysSerValThrGlnGlyLeuAlaTyrAspArgAsnAsnArgMetSerLeu 569  
 QY 520 CTCGTATCGTAGACCTTTCTATTCCAAACGAAGTAATCTTCTCTCATCGTCTACGATGTA 579  
 DB 570 TyrGlyValLeuHisAlaLysTyrGlySerTrpAspPheTyr-----583  
 QY 580 TTCAAGGAGGTTTCGTATGGTATTTGCTCTCTCCAGCTCTAGTAGTAACTTCGAGGCGCAT 639  
 DB 584 -----AsnTyrGlyPheAla 588  
 QY 640 ACGGACAACTGATGTGCGCGCTCACACGCGGCGACTTCAGCGTATTCGCGGTGTATGCC 699  
 DB 599 LeuSerAsnTyrMetTyrAsnAsnMetGlyMetPhe-----601  
 QY 700 GGTCCGACAAACCGCGCGCGAATACAGCAAGGACAATAACCCCTATAAGCCCGTTTAC 759  
 DB 602 -----AsnLysMetThrAsnTyrIleLysAsnAsn-----611  
 QY 760 TTCGCTGCGGTATCCATCAAGGCTACAAG-----GCTGACGACTAT 801  
 DB 612 -----AspValSerGlyTyrLysAspTyrIleAlaSerMetSerSerAspTyr 627  
 QY 802 GCCATGACCATCGGTTTCCCGGCGAGTACGAGTACCTACCTCCTCTTGGGGTGTGAA 861  
 DB 628 GlyLeuAsnAspLysTyrGlnAspTyrMetAspSerLeuLeu-----641  
 QY 862 GATCGTATCGAAACGAGAAACAACTCTCGTATCGAA-----GTTCCGGGTATCAAG 912  
 DB 862 GATCGTATCGAAACGAGAAACAACTCTCGTATCGAA-----GTTCCGGGTATCAAG 912

```
Db 642 AsnAsnIleAspAsnLeuValProLeuValSerAspGluTyrValAsnGly----- 659
QY 913 CAAGCATCTGGAAGAACCCATGAGCGGAGATCAGGCTACCCGT---ATCAAAATATGCC 969
Db 660 -----HisGluAlaLysAspIleAsnGluIleThrAsnAspIleLysGluVal 675
QY 970 AGCAAGATGCTCAG---AGTGCTAACTATTGGAAGAATTTCGATCGGTATGAACCGCGT 1026
Db 676 SerAsnIleLysAspLeuSerAsnValGluLysSerGlnPhePheThrThrTyrAsp 695
QY 1027 CTCGCTGCTCTGAGTGTAGTGTAGCGTGTAGCGTGTAGCGGAGGAAGACATTCGCAGACTGG 1086
Db 696 MetArgGlyThrTyrValGlyGlyArgSerGlnGlyGluAsn-----AspTrp 712
QY 1087 ATCCGTAAGACGCGAAGAGTCTGCTATGCGGATGTTCTGCTCTCTCCTCAAAAG--- 1143
Db 713 LysAspMetAsnSerLys-----LeuAsnAspIleLeuLysGluLeuSerLysLys 729
QY 1144 -----GCTTATAAGGAAGAGCCCAAGCCCAACCGTGAG 1176
Db 730 SerTrpAsnGlyTyrLysThrValThrAlaTyrPheValAsnHisLysValAspGlyAsn 749
QY 1177 ATGACTTATTTCAGGAGAGCGCTCTTCGGTGTGTCGAGGTGGTTCGTTTCCACAGATT 1236
Db 750 GlyAsnTyrValTyrAspValValPheHisGlyMetAsnThr----- 763
QY 1237 GCCAAGCATTCGCTACAAATCTCATGCTCATGCC-----GGTATCCTC 1281
Db 764 -----AspThrAsnThrAspValHisValAsnLysGluProLysAlaValIle 779
QY 1282 AAATCGCTTGACGACAAGTACAAGACTACCTCCCTCGCTCGCTGAAGGTGCTGCC 1341
Db 780 LysSer-----AspSerSerVal----- 785
QY 1342 GCCATCTCGATATTGTAGCGCGGGGTATCCCTCCGCGACAAGCTCCCGATATATTCAG 1401
Db 786 -----IleValGluGluGluIleAsn-----PheAsp 794
QY 1402 AATGTAATCGACAAGAAATTCAAAGCGGACAGCAAGAAGTAT----- 1443
Db 795 GlyThrGluSerLysAspGluAspGlyGluIleLysAlaTyrGluTrpAspPheGlyAsp 814
QY 1444 -----CGAGACTTCGTATTTCGACAAGAGTGTGTTCTCTTAT 1479
Db 815 GlyGluLysSerAsnGluAlaLysAlaThrHisLysTyrAsnLys----- 829
QY 1480 ACGGACAAGTTCATCGCATGCTCAAGTCCATGGAC-----AAG 1518
Db 830 ThrGlyGluTyrGluValLysLeuThrValThrAspAsnAsnGlyGlyIleAsnThrGlu 849
QY 1519 GAAAAGTTTGCACAGGCTATCGAGAAAGATCCG----- 1551
Db 850 SerLysLysIleLysValValGluAspLysProValGluValIleAsnGluSerGluPro 869
QY 1552 -----CGACTAGAGCTTTCACAGAGCGTAAAGCTGCTCGCTCGCGCT 1593
Db 870 AsnAsnAspPheGluLysAlaAsnGlnIleAlaLysSerAsnMetLeuValLysGlyThr 889
QY 1594 ATTCAGCGCGATCGCATCGCCATCCCTATGCCATTGAGAGGGAACGGT----- 1644
Db 890 LeuSerGluGluAspTyrSerAspLysTyrTyrPheAspValAlaLysLysGlyAsnVal 909
QY 1645 -----CTTTTCTTTCCCGGT 1659
Db 910 LysIleThrLeuAsnAsnLeuAsnSerValGlyIleThrTrpThrLeuTyrLysGluGly 929
QY 1660 TTGCTGTAGATGTAC-----CCCGCAGTGTCTGCTCCGAGCGAT 1698
Db 930 AspLeuAsnAsnTyrValLeuTyrAlaThrGlyAsnAspGlyThrValLeuLysGlyGlu 949
QY 1699 GCCAATCTCACCATCGGTATGAGCTACGGCTCCATCAGGAGGATGATGACCGCAGACCGT 1758
Db 950 LysThrLeuGluProGlyArgTyrTyrLeuSerValTyrThrTyrAspAsnGlnSerGly 969
```

```
QY 1759 GCCTGGTACACATATCATACGACAGCGGCGGTATTGGAGAACGAGATCCTAAGACC 1818
Db 970 -----ThrTyrThrValAsnValLysGlyAsnLeuLysAsnGluValLysGluThr 986
QY 1819 GATGAGTTTGGCGGTACAGGAGAATAATCTCGACCTCTTCCGCAACCAAACTATGCTCGC 1878
Db 987 AlaLysAspAlaIleLysGlu----- 993
QY 1879 TATGCCGGAAGAGGTCAGCTCCATATCGCTTCTTCATCGAACACGACATCACGGCGGT 1938
Db 994 ---ValGluAsn-----AsnAsnAsp----- 999
QY 1939 RACTCCGGTAGCCCGCTATTCGATAAG-----AACGGCGCTCTGATC 1980
Db 1000 -----PheAspLysAlaMetLysValAspSerAsnSerLysIleVal 1013
QY 1981 GGT---CTTGCTTTTCGATGGCAACTGGGAAGCTATGAGTGGTGCATCGACTTCCAACCC 2037
Db 1014 GlyThrLeuSerAsnAspAspLeuLysAspIleTyrSerIleAspIleGlnAsnProSer 1033
QY 2038 GATCTGACGCGCACAACTC-----ACGTGGGACATCCGCTAGCTTCTCTTCATG 2085
Db 1034 AspLeuAsnIleValIleValGluAsnLeuAspAsnIleLysMetAsnTrpLeuLeuTyrSer 1053
QY 2086 ATTGAC 2091
Db 1054 AlaAsp 1055

RESULT 11
NAGH_CLOPE STANDARD; PRT; 1628 AA.
AC P26831;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hyaluronoglucosaminidase precursor (EC 3.2.1.35) (Hyaluronidase)
DE (Mu toxin).
GN NAGH OR CPE0191.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50.
RX MEDLINE=94232189; PubMed=8177218;
RA Canard B., Garnier T., Saint-Joanis B., Cole S.T.;
RT "Molecular genetic analysis of the nagH gene encoding a hyaluronidase
of Clostridium perfringens.";
RL Mol. Gen. Genet. 243:215-224(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -!- FUNCTION: Putative virulence factor which is likely to act on
connective tissue during gas gangrene.
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-linkages between N-
acetyl-beta-D-glucosamine and D-glucuronate residues in
hyaluronate.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- CAUTION: The partially purified protein from strain CPN50 is
approximately 70 kDa smaller than the sequence indicated here.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
```

modified and this statement is not removed. Usage by and for communit  
entities requires a license agreement (see http://www.isb-sib.ch/annou  
or send an email to license@isb-sib.ch).

EMBL; M81878; AAA233259.1; -  
ENBL; AP003185; BAB79897.1; -.  
PIR; S27540; S27540.  
InterPro; IPR002105; Dockerin\_1.  
InterPro; IPR000421; FA58\_C.  
Pfam; PF00404; Dockerin\_1; 2.  
Pfam; PF00754; F5\_F8\_type\_C; 1.  
SMART; SM00231; FA58C; 1.  
Hydrolase; Glycosidase; Toxin; Repeat; Signal; Complete proteome.

FT SIGNAL 1 30 POTENTIAL.  
FT CHAIN 31 1628 HYALURONOGUCOSAMINIDASE.  
FT DOMAIN 703 1032 3 X APPROXIMATE REPEATS.  
FT REPEAT 703 748 1.  
FT REPEAT 848 894 2.  
FT REPEAT 987 1032 3.  
FT VARIANT 147 147 G -> A (IN STRAIN CPN50).  
FT VARIANT 172 175 KIQS -> EIKN (IN STRAIN CPN50).  
FT VARIANT 250 250 V -> M (IN STRAIN CPN50).  
FT VARIANT 548 548 A -> E (IN STRAIN CPN50).  
FT VARIANT 558 558 D -> E (IN STRAIN CPN50).  
FT VARIANT 614 614 G -> S (IN STRAIN CPN50).  
FT VARIANT 944 944 I -> V (IN STRAIN CPN50).  
FT VARIANT 950 950 N -> S (IN STRAIN CPN50).  
FT VARIANT 979 979 T -> I (IN STRAIN CPN50).  
FT VARIANT 982 982 I -> L (IN STRAIN CPN50).  
FT VARIANT 1042 1042 I -> F (IN STRAIN CPN50).  
FT VARIANT 1043 1628 MISSING (IN STRAIN CPN50).  
SQ SEQUENCE 1628 AA; 182474 MW; D4252A2512BBED69 CRC64;

Alignment Scores:  
Pred. No.: 2,75 Length: 1628  
Score: 121.00 Matches: 129  
Percent Similarity: 32.24% Conservative: 78  
Best Local Similarity: 20.09% Mismatches: 209  
Query Match: 3.17% Indels: 226  
DB: 1 Gaps: 36

US-10-008-355-1 (1-2139) x NAGH\_CLOPE (1-1628)

Qy 526 ATCGTAGAGCTTTTCATTATCCAAAC-----GAATACTTCTCATCGTCTAC 573  
::|::|:: |::|::|:: ||||| |||||  
Db 111 ValValAsnThrArgPheAsnLysAsnIleProHisAspGluSerPhe-----Phe 127  
  
Qy 574 GATGTATTCAAGGACGTCGTGTGGTATTTGCCTCCGACGTCCTAGTAGTTCGGA 633  
||||| |||||  
Db 128 AspGluLysMetAspAlaAsnIleValSerValLysAspGlyValIleGlyValIleGly 147  
  
Qy 634 GGCGATACGGAACAACGATGTGGCCGCGCACACGCCGACTTCAGCGCTATTCCGCGTG 693  
||||| |||||  
Db 148 GluAspThrAspSerAlaPheTyrr-----GlyValThrThrLeuLysHisVal 163  
  
Qy 694 TATCCGGTCGCCAACACC GCCGCGCGCAATACAGAAGCAACAATAAACCCCTATAAGCCC 753  
::| ::|::|::| |||||  
Db 164 Phe-----AsnGlnLeuGluGluGlyAsnLys----- 172  
  
Qy 754 GTTTACTTCGTCGCGTATCCATGCGGCTACAGCGCTACAGGACATATGCGATGACCATC 813  
::|::|::| :|||:|::|::|::|::|::|::|  
Db 173 -----IleGlnSerPheArgAlaAspTyrAla---GluVal 184  
  
Qy 814 GGTTTCCCCGAGCTACGGATCGCTACCTC-- --ACTTCTTGGGGTGTGGAAGATCGTATC 870  
||||| |||||  
Db 185 AlaHisArgGlyPheIleGluGlyTyrrGlyAsnProTrpSerAsnGluaspArgAla 204  
  
Qy 871 GAA-----AACGAGAAC 882  
|||  
Db 205 GluLeuMetLysPheGlyGlyAspTyrLysLeuAsnGlnTyrValPheAlaProLysAsp 224  
  
Qy 883 AATCCTCGTATCGAAGTTCCGGGTATCAAGCAAGGCGATCTCGAAGGAAGCCATGACGCA 942  
::|::|::| |||||

```

Qy 1720 AGTACGGCTCCATCAAGGGA-----TATGAACCGCAGCAGCGTCCCTGG 1764
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 587 AlalileGlyThrLeuLysSerAlaIleAlaIleGluGluGlyAspGluAlaAlaTrp 606
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 1765 TACAACATATACACAGCAGGCAAGGCGCTATTGGAGAACAGCAGGATCCTTAAGAGCGATGAG 1824
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 607 AlaAsnTyr---SerGluAlaGlnGlyAlaPheGluLys----- 618
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 1825 TTGCCGTACAGGAGATATCTCGACCTCTCCGCCACCAAAACTATGGT----- 1875
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 619 -----SerLysThrTyrGlyPheHisTyr 626
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 1876 -----CGCTATGCGGAGACGCTCAGCTCCATATC----- 1905
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 627 ValAspHisThrGluTyrAlaGluValGlyValGlnHisIleValPropheIleLysSer 646
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 1905 ----- 1905
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 647 MetGlyGlnAsnLeuSerValValIleGlySerIleValAspProAsnArgIleAla 666
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 1906 GCTTTCCTATCGAAC---AAGCAGATCAGCGGCTACTCGGTAGCCCGCTATTCGAT 1962
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 667 ThrTyrIleSerAsnArgGlnAspAlaProThrGlyAsnProAspAsn---IlePheAsp 685
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 1963 AAGAAC 1968
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 686 AsnAsn 587
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::

RESULT 12
ID EFG_VIBCH STANDARD; PRT; 698 AA.
AC O9KU27;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Elongation factor G (EF-G).
GN FUSA OR VC0361.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID-666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- FUNCTION: This protein promotes the GTP-dependent translocation of
CC the nascent protein chain from the A-site to the P-site of the
CC ribosome.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE004124; AAF93534.1;
CC HSSP; P13551; IELO.
CC TIGR; VC0361;

```

```

DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFT.
DR TIGRFAMS; TIGR00484; EF-G; 1.
DR TIGRFAMS; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding.
KW Complete proteome.
FT NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 88 92 GTP (BY SIMILARITY).
FT NP_BIND 142 145 GTP (BY SIMILARITY).
SQ SEQUENCE 698 AA; 76927 MW; F44FDA4DB8FC4ECC CRC64;

```

## Alignment Scores:

```

Pred. No.: 3.3 Length: 698
Score: 118.00 Matches: 120
Percent Similarity: 31.18% Conservative: 83
Best Local Similarity: 18.43% Mismatches: 202
Query Match: 3.09% Indels: 246
DB: 1 Gaps: 26

```

US-10-008-355-1 (1-2139) x EFG\_VIBCH (1-698)

```

Qy 22 ATTCTTCTCGAGCAGCCCTGCTGTTG---GGTGCTTCAGGGTAGCCAAAGCCGACAAA 78
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 105 ValLeuAspGlyAlaValAlaValPheCysGlyThrSerGlyValGluProGlnSerGlu 124
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 79 GGCATGTGG-----CTCTCAACGACATC 102
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 125 ThrValTrpArgGlnAlaAspLysTyrGlyValProArgMetValPheValAsnLysMet 144
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 103 AATCAGGAGAAATCTGGATCGGAATCGGTGAGCTCGGTTCACGCTCCCGTGGATCGCTC 162
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 145 AspArgAlaGlyAlaAspPheLeuArgValGly----- 156
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 163 TACAGTTTCGACAAAGCCGTCCTCCATTGCCAATCCGCTGTTATC----- 204
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 157 ---GlnIleLysHisArgLeuGlyAlaAsnProValProIleGlnLeuAsnIleGlyAla 175
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 205 -----FTCCGTGCGGGATGTACCGGTATCAGTG----- 234
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 176 GluGluGluPheLysGlyValIleAspLeuIleLysMetLysAlaIleAsnTrpAsnGlu 195
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 235 TCCGATCAGGCGCTGATCTTTTACC----- 258
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 196 AlaAspGlnGlyMetSerPheThrTyrGluIleProAlaAspMetLeuGluLeuAla 215
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 259 -----AACCCACCTCGCGATACGCTGCTATCCAGACGCAAGCCGATGATCAC 309
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 216 GlnGluTrpArgAsnHisLeuValGluAlaAlaGluAlaSerGluGluLeuMetGlu 235
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 310 GACTATCTGCCGATGGTTTCGTTCT-----CGC 339
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 236 LysTyrLeuGluAspGlyGluLeuSerGluValGluIleLysGlnAlaLeuArgGlnArg 255
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 340 ACGATGGGTGAGGATCTCCGATTCCTCCGCTGTTCCGTTGAGTATCTGCCGCAAGCATGCTG 399
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 256 ThrIleAsnGlnIleValLeuAlaAlaCysGlySerAlaPheLysAsnLysGlyVal 275
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 400 AAG---GTAACGGCAAGGTA-----GAAGGACAGCTC 429
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 276 GlnAlaValLeuAspAlaValIleGluPheLeuProSerProThrAspValProAlaIle 295
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Qy 430 AAGGGTATCACTGACGAGATGGACGCTCTCGCAAAAGCTCAGGAGGTATGCCAAGAAGTC 489
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db 296 LysGlyIleAspAsp-----ArgGluAsn 303
    ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::

```

QY 490 GCCAAAGAAATCGACGAGCAACCACTCTGCATCGTAGAGCCTTCTATTCCAC 549  
Db 304 SerValGluArgHisAlaAspAsn-----GluProPhe-----Ser 316  
QY 550 AAGCAATACCTTCTCATCGCTACGATGATTCAGGACGCTTCGTATGTTGCTCCT 609  
Db 317 SerLeuAlaPheLysIleAlaThrAspPhe----- 327  
QY 610 CCCAGCTCTGTAGGTAACTTGGAGGCGATACGACAACTGGATGGCCCGCTACACG 669  
Db 328 -----Val 328  
QY 670 GCGCACTTACGCTATTCCTCGGTGTATCGGTGCCGACCAACCGCCGCGGATACAGC 729  
Db 329 GlySerLeuThrPheIleArgValTySerGlyValValAsnSerGlyAspAlaValTy 348  
QY 730 AAGGACAATAAACCCCTATAAGCTTACTTCTCGCTCCGCTATCATCAAGGCTACAAG 789  
Db 349 AsnSerValLysGlnLysLysGluArgPheGlyArgIleValGlnMetHisAlaAsnLys 368  
QY 790 GCTGAC-----GACTATGCCATGACCATCGGT----- 816  
Db 369 ArgAspGluIleLysGluIleArgAlaGlyAspIleAlaAlaIleGlyLeuLysAsp 388  
QY 816 ----- 816  
Db 389 ValThrThrGlyAspThrLeuLysAspProAsnHisValValIleLeuGluArgMetGlu 408  
QY 817 TTCGCGGCGAGTACGGATCGCTACCTCCTCTTGGGTGTGGAAGATCGTATCGAAAC 876  
Db 409 PhePro-----GluProValIleGlnIle 416  
QY 877 GAGAACAACTCTCTATPCGAAGTTCGGGTATCAAGCAAGCATCTGGAAGAGCCATG 936  
Db 417 AlaValGluProArgSerLysAlaAspGlnLysMetGlyIleAlaLeuGlyLysLeu 436  
QY 937 AGCGCA--GATCAGGTACCGCTATCAATATGCCAGCAAGTATGCTCAGAT----- 987  
Db 437 AlaAlaGluAspProSerPheArgValGluThrAspAlaGluThrGlyGlnThrLeuIle 456  
QY 988 -----GCTAACTATTGGAAGAAATTCGATCGGT 1014  
Db 457 SerGlyMetGlyGluLeuHisLeuAspIleValAspArgMetLysArgGluPheGly 476  
QY 1015 ATGAACCGCGTCTCGCTCTTTCAGCTGATAGTCTGAAG-----CGTCCGCGAGAA 1068  
Db 477 ValAspCysAsnValGlyLysProGlnValAlaTyArgGluThrIleArgGlyLysSer 496  
QY 1069 AGACATTCGACACTGGATCCGTAAGCAAGCGGAGAGTCTGCTATGCG----- 1119  
Db 497 GluValGluGlyLysPheValArgGlnSerGlyGlyArgGlyGlnTyGlyHisValTrp 516  
QY 1119 ----- 1119  
Db 517 LeuLysIleGluProAlaGluProGlyGlnGlyPheValPheValAspAlaIleAlaGly 536  
QY 1120 -----GATGTAFTGTCTTCTCTCGAAAGCGTTATAAG 1152  
Db 537 GlyValIleProLysGluPheIleAsnProValAlaLysGlyIleGluGlnMetAsn 556  
QY 1153 GAAGGCCCAAGCCACCGTGAGTACTATTTGAGCAGACGCTCTCGGTGGTACC 1212  
Db 557 AsnGlyValLeuAlaGlyTyProValLeuAspValLysAlaThrLeuPheAspGlySer 576  
QY 1213 ---GAGTGGTCTGTTTTCACAGTTTCCCAACGCAATGCTGCTACAAATCCTGATGCTCAT 1269  
Db 577 PheHisAspValAspSerSerGluMetAlaPheLysIleAlaGlySerMetAlaPheLys 596  
QY 1270 GCCGGTATCCTCAATCGCTT----- 1290  
Db 597 LysGlyAlaLeuGluAlaGlnProValLeuGluProLeuMetLysValGluIleThr 616  
QY 1291 -----GACGACAAGTACAAGACTACTCCCTCGCTCGACCGTAAG-----GTCTGT 1338

Db 617 ThrProGluAspTrpMetGlyAspValValGlyAspLeuAsnArgArgGlyIleIle 636  
QY 1339 -----CCGCGCATCTCGATATTTACGCGCGGTATCCCTGCGGAC 1380  
Db 637 GluGlyMetAspGluGlyProAlaGlyLeuLysIleIleHisAlaLysValPro----- 654  
QY 1381 AAGCTCCCCGATATATTCAGAATGTATATCGACAAGAA----- 1419  
Db 655 ---LeuSerGluMetPheGlyTyAlaThrAspLeuArgSerAlaThrGlnGlyArgAla 673  
QY 1420 ---TTCAAGCGGCACACAGAAAGTATGTCAGAC 1449  
Db 674 SerTySerMetGluPheAlaGluTyAlaAsp 684  
RESULT 13  
HLVA\_ECOLI  
ID HLVA\_ECOLI STANDARD; PRT; 1024 AA.  
AC P08715;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hemolysin, plasmid.  
GN HLVA.  
OS Escherichia coli.  
OG Plasmid Inci2 pHLy152.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hess J., Wells W., Vogel M., Goebel W.;  
RT "Nucleotide sequence of a plasmid-encoded hemolysin determinant and  
its comparison with a corresponding chromosomal hemolysin sequence.";  
RL FEMS-Microbiol. Lett. 34:1-11(1986).  
RN [2]  
RP PALMITOYLATION OF LYS-564 AND LYS-690.  
RX MEDLINE=95099325; PubMed=7801126;  
RA Stanley P., Packman L.C., Koronakis V., Hughes C.;  
RT "Fatty acylation of two internal lysine residues required for the  
toxic activity of Escherichia coli hemolysin.";  
RL Science 266:1992-1996(1994).  
RN [3]  
RP PALMITOYLATION OF LYS-564 AND LYS-690.  
RX MEDLINE=96404790; PubMed=8808931;  
RA Ludwig A., Garcia F., Jarchau T., Benz R., Hoppe J.,  
Goebel W.;  
RT "Analysis of the in vivo activation of hemolysin (HlyA) from  
Escherichia coli.";  
RL J. Bacteriol. 178:5422-5430(1996).  
CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
DEFINED.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING  
CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC  
ACTIVITY.  
CC -!- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE  
INVOLVED IN PORE FORMATION BY THE CYTOTOXIN.  
CC -!- PTM: PALMITOYLATED BY HLYC. THE TOXIN ONLY BECOMES ACTIVE WHEN  
MODIFIED.  
CC -!- DISEASE: THE HEMOLYSIN OF E.COLI IS PRODUCED PREDOMINANTLY BY  
STRAINS CAUSING EXTRAINTESTINAL INFECTIONS, SUCH AS THOSE OF THE  
URINARY TRACT.  
CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

```

CC -----
DR EMBL; M14107; AAA98233.1; ..
DR InterPro; IPR001343; Hemolysn_Ca_bind.
DR InterPro; IPR003355; RTXtoxin_N.
DR InterPro; IPR003995; RtxA.
DR Pfam; PF00353; hemolysinCbind; 6.
DR Pfam; PF02382; RTX; 1.
DR PRINTS; PRO0313; CABNDNGRPT.
DR PRINTS; PRO1488; RTXTOXINA.
DR PROSITE; PS00330; HEMOLYSIN_CALCIIUM; 4.
KW Hemolysis; Toxin; Cytolysis; Cytotoxin; Repeat; Calcium;
KW Transmembrane; Lipoprotein; Palmitate; Plasmid.
FT TRANSMEM 238 260
FT TRANSMEM 268 327
FT TRANSMEM 365 411
FT DOMAIN 724 870
FT REPEAT 724 729
FT REPEAT 733 738
FT REPEAT 742 747
FT REPEAT 751 756
FT REPEAT 760 765
FT REPEAT 769 774
FT REPEAT 778 783
FT REPEAT 787 792
FT REPEAT 796 801
FT REPEAT 807 812
FT REPEAT 817 822
FT REPEAT 826 831
FT REPEAT 835 840
FT REPEAT 844 849
FT REPEAT 856 861
FT REPEAT 865 870
FT REPEAT 879 884
FT REPEAT 893 898
FT REPEAT 907 912
FT REPEAT 916 921
FT REPEAT 925 930
FT REPEAT 934 939
FT REPEAT 943 948
FT REPEAT 952 957
FT REPEAT 961 966
FT REPEAT 969 974
FT REPEAT 978 983
FT REPEAT 987 992
FT REPEAT 996 1001
FT REPEAT 1005 1010
FT REPEAT 1014 1019
FT REPEAT 1023 1028
FT REPEAT 1032 1037
FT REPEAT 1041 1046
FT REPEAT 1049 1054
FT REPEAT 1058 1063
FT REPEAT 1067 1072
FT REPEAT 1076 1081
FT REPEAT 1085 1090
FT REPEAT 1094 1099
FT REPEAT 1103 1108
FT REPEAT 1112 1117
FT REPEAT 1121 1126
FT REPEAT 1130 1135
FT REPEAT 1139 1144
FT REPEAT 1148 1153
FT REPEAT 1157 1162
FT REPEAT 1166 1171
FT REPEAT 1175 1180
FT REPEAT 1184 1189
FT REPEAT 1193 1198
FT REPEAT 1202 1207
FT REPEAT 1211 1216
FT REPEAT 1220 1225
FT REPEAT 1229 1234
FT REPEAT 1238 1243
FT REPEAT 1247 1252
FT REPEAT 1256 1261
FT REPEAT 1265 1270
FT REPEAT 1274 1279
FT REPEAT 1283 1288
FT REPEAT 1292 1297
FT REPEAT 1301 1306
FT REPEAT 1310 1315
FT REPEAT 1319 1324
FT REPEAT 1328 1333
FT REPEAT 1337 1342
FT REPEAT 1346 1351
FT REPEAT 1355 1360
FT REPEAT 1364 1369
FT REPEAT 1373 1378
FT REPEAT 1382 1387
FT REPEAT 1391 1396
FT REPEAT 1400 1405
FT REPEAT 1409 1414
FT REPEAT 1418 1423
FT REPEAT 1427 1432
FT REPEAT 1436 1441
FT REPEAT 1445 1450
FT REPEAT 1454 1459
FT REPEAT 1463 1468
FT REPEAT 1472 1477
FT REPEAT 1481 1486
FT REPEAT 1490 1495
FT REPEAT 1500 1505
FT REPEAT 1509 1514
FT REPEAT 1518 1523
FT REPEAT 1527 1532
FT REPEAT 1536 1541
FT REPEAT 1545 1550
FT REPEAT 1554 1559
FT REPEAT 1563 1568
FT REPEAT 1572 1577
FT REPEAT 1581 1586
FT REPEAT 1590 1595
FT REPEAT 1600 1605
FT REPEAT 1609 1614
FT REPEAT 1618 1623
FT REPEAT 1627 1632
FT REPEAT 1636 1641
FT REPEAT 1645 1650
FT REPEAT 1654 1659
FT REPEAT 1663 1668
FT REPEAT 1667 1672
FT REPEAT 1671 1676
FT REPEAT 1675 1680
FT REPEAT 1679 1684
FT REPEAT 1683 1688
FT REPEAT 1687 1692
FT REPEAT 1691 1696
FT REPEAT 1695 1700
FT REPEAT 1709 1714
FT REPEAT 1718 1723
FT REPEAT 1727 1732
FT REPEAT 1736 1741
FT REPEAT 1745 1750
FT REPEAT 1754 1759
FT REPEAT 1763 1768
FT REPEAT 1772 1777
FT REPEAT 1781 1786
FT REPEAT 1790 1795
FT REPEAT 1800 1805
FT REPEAT 1809 1814
FT REPEAT 1813 1818
FT REPEAT 1817 1822
FT REPEAT 1821 1826
FT REPEAT 1825 1830
FT REPEAT 1829 1834
FT REPEAT 1833 1838
FT REPEAT 1837 1842
FT REPEAT 1841 1846
FT REPEAT 1845 1850
FT REPEAT 1849 1854
FT REPEAT 1853 1858
FT REPEAT 1857 1862
FT REPEAT 1861 1866
FT REPEAT 1865 1870
FT REPEAT 1869 1874
FT REPEAT 1873 1878
FT REPEAT 1877 1882
FT REPEAT 1881 1886
FT REPEAT 1885 1890
FT REPEAT 1889 1894
FT REPEAT 1893 1898
FT REPEAT 1897 1902
FT REPEAT 1901 1906
FT REPEAT 1905 1910
FT REPEAT 1909 1914
FT REPEAT 1913 1918
FT REPEAT 1917 1922
FT REPEAT 1921 1926
FT REPEAT 1925 1930
FT REPEAT 1929 1934
FT REPEAT 1933 1938
FT REPEAT 1937 1942
FT REPEAT 1941 1946
FT REPEAT 1945 1950
FT REPEAT 1949 1954
FT REPEAT 1953 1958
FT REPEAT 1957 1962
FT REPEAT 1961 1966
FT REPEAT 1965 1970
FT REPEAT 1969 1974
FT REPEAT 1973 1978
FT REPEAT 1977 1982
FT REPEAT 1981 1986
FT REPEAT 1985 1990
FT REPEAT 1989 1994
FT REPEAT 1993 1998
FT REPEAT 1997 2002
FT REPEAT 2001 2006
FT REPEAT 2005 2010
FT REPEAT 2009 2014
FT REPEAT 2013 2018
FT REPEAT 2017 2022
FT REPEAT 2021 2026
FT REPEAT 2025 2030
FT REPEAT 2029 2034
FT REPEAT 2033 2038
FT REPEAT 2037 2042
FT REPEAT 2041 2046
FT REPEAT 2045 2050
FT REPEAT 2049 2054
FT REPEAT 2053 2058
FT REPEAT 2057 2062
FT REPEAT 2061 2066
FT REPEAT 2065 2070
FT REPEAT 2069 2074
FT REPEAT 2073 2078
FT REPEAT 2077 2082
FT REPEAT 2081 2086
FT REPEAT 2085 2090
FT REPEAT 2089 2094
FT REPEAT 2093 2098
FT REPEAT 2097 2102
FT REPEAT 2101 2106
FT REPEAT 2105 2110
FT REPEAT 2109 2114
FT REPEAT 2113 2118
FT REPEAT 2117 2122
FT REPEAT 2121 2126
FT REPEAT 2125 2130
FT REPEAT 2129 2134
FT REPEAT 2133 2138
FT REPEAT 2137 2142
FT REPEAT 2141 2146
FT REPEAT 2145 2150
FT REPEAT 2149 2154
FT REPEAT 2153 2158
FT REPEAT 2157 2162
FT REPEAT 2161 2166
FT REPEAT 2165 2170
FT REPEAT 2169 2174
FT REPEAT 2173 2178
FT REPEAT 2177 2182
FT REPEAT 2181 2186
FT REPEAT 2185 2190
FT REPEAT 2189 2194
FT REPEAT 2193 2198
FT REPEAT 2197 2202
FT REPEAT 2201 2206
FT REPEAT 2205 2210
FT REPEAT 2209 2214
FT REPEAT 2213 2218
FT REPEAT 2217 2222
FT REPEAT 2221 2226
FT REPEAT 2225 2230
FT REPEAT 2229 2234
FT REPEAT 2233 2238
FT REPEAT 2237 2242
FT REPEAT 2241 2246
FT REPEAT 2245 2250
FT REPEAT 2249 2254
FT REPEAT 2253 2258
FT REPEAT 2257 2262
FT REPEAT 2261 2266
FT REPEAT 2265 2270
FT REPEAT 2269 2274
FT REPEAT 2273 2278
FT REPEAT 2277 2282
FT REPEAT 2281 2286
FT REPEAT 2285 2290
FT REPEAT 2289 2294
FT REPEAT 2293 2298
FT REPEAT 2297 2302
FT REPEAT 2301 2306
FT REPEAT 2305 2310
FT REPEAT 2309 2314
FT REPEAT 2313 2318
FT REPEAT 2317 2322
FT REPEAT 2321 2326
FT REPEAT 2325 2330
FT REPEAT 2329 2334
FT REPEAT 2333 2338
FT REPEAT 2337 2342
FT REPEAT 2341 2346
FT REPEAT 2345 2350
FT REPEAT 2349 2354
FT REPEAT 2353 2358
FT REPEAT 2357 2362
FT REPEAT 2361 2366
FT REPEAT 2365 2370
FT REPEAT 2369 2374
FT REPEAT 2373 2378
FT REPEAT 2377 2382
FT REPEAT 2381 2386
FT REPEAT 2385 2390
FT REPEAT 2389 2394
FT REPEAT 2393 2398
FT REPEAT 2397 2402
FT REPEAT 2401 2406
FT REPEAT 2405 2410
FT REPEAT 2409 2414
FT REPEAT 2413 2418
FT REPEAT 2417 2422
FT REPEAT 2421 2426
FT REPEAT 2425 2430
FT REPEAT 2429 2434
FT REPEAT 2433 2438
FT REPEAT 2437 2442
FT REPEAT 2441 2446
FT REPEAT 2445 2450
FT REPEAT 2449 2454
FT REPEAT 2453 2458
FT REPEAT 2457 2462
FT REPEAT 2461 2466
FT REPEAT 2465 2470
FT REPEAT 2469 2474
FT REPEAT 2473 2478
FT REPEAT 2477 2482
FT REPEAT 2481 2486
FT REPEAT 2485 2490
FT REPEAT 2489 2494
FT REPEAT 2493 2498
FT REPEAT 2497 2502
FT REPEAT 2501 2506
FT REPEAT 2505 2510
FT REPEAT 2509 2514
FT REPEAT 2513 2518
FT REPEAT 2517 2522
FT REPEAT 2521 2526
FT REPEAT 2525 2530
FT REPEAT 2529 2534
FT REPEAT 2533 2538
FT REPEAT 2537 2542
FT REPEAT 2541 2546
FT REPEAT 2545 2550
FT REPEAT 2549 2554
FT REPEAT 2553 2558
FT REPEAT 2557 2562
FT REPEAT 2561 2566
FT REPEAT 2565 2570
FT REPEAT 2569 2574
FT REPEAT 2573 2578
FT REPEAT 2577 2582
FT REPEAT 2581 2586
FT REPEAT 2585 2590
FT REPEAT 2589 2594
FT REPEAT 2593 2598
FT REPEAT 2597 2602
FT REPEAT 2601 2606
FT REPEAT 2605 2610
FT REPEAT 2609 2614
FT REPEAT 2613 2618
FT REPEAT 2617 2622
FT REPEAT 2621 2626
FT REPEAT 2625 2630
FT REPEAT 2629 2634
FT REPEAT 2633 2638
FT REPEAT 2637 2642
FT REPEAT 2641 2646
FT REPEAT 2645 2650
FT REPEAT 2649 2654
FT REPEAT 2653 2658
FT REPEAT 2657 2662
FT REPEAT 2661 2666
FT REPEAT 2665 2670
FT REPEAT 2669 2674
FT REPEAT 2673 2678
FT REPEAT 2677 2682
FT REPEAT 2681 2686
FT REPEAT 2685 2690
FT REPEAT 2689 2694
FT REPEAT 2693 2698
FT REPEAT 2697 2702
FT REPEAT 2701 2706
FT REPEAT 2705 2710
FT REPEAT 2709 2714
FT REPEAT 2713 2718
FT REPEAT 2717 2722
FT REPEAT 2721 2726
FT REPEAT 2725 2730
FT REPEAT 2729 2734
FT REPEAT 2733 2738
FT REPEAT 2737 2742
FT REPEAT 2741 2746
FT REPEAT 2745 2750
FT REPEAT 2749 2754
FT REPEAT 2753 2758
FT REPEAT 2757 2762
FT REPEAT 2761 2766
FT REPEAT 2765 2770
FT REPEAT 2769 2774
FT REPEAT 2773 2778
FT REPEAT 2777 2782
FT REPEAT 2781 2786
FT REPEAT 2785 2790
FT REPEAT 2789 2794
FT REPEAT 2793 2798
FT REPEAT 2797 2802
FT REPEAT 2801 2806
FT REPEAT 2805 2810
FT REPEAT 2809 2814
FT REPEAT 2813 2818
FT REPEAT 2817 2822
FT REPEAT 2821 2826
FT REPEAT 2825 2830
FT REPEAT 2829 2834
FT REPEAT 2833 2838
FT REPEAT 2837 2842
FT REPEAT 2841 2846
FT REPEAT 2845 2850
FT REPEAT 2849 2854
FT REPEAT 2853 2858
FT REPEAT 2857 2862
FT REPEAT 2861 2866
FT REPEAT 2865 2870
FT REPEAT 2869 2874
FT REPEAT 2873 2878
FT REPEAT 2877 2882
FT REPEAT 2881 2886
FT REPEAT 2885 2890
FT REPEAT 2889 2894
FT REPEAT 2893 2898
FT REPEAT 2897 2902
FT REPEAT 2901 2906
FT REPEAT 2905 2910
FT REPEAT 2909 2914
FT REPEAT 2913 2918
FT REPEAT 2917 2922
FT REPEAT 2921 2926
FT REPEAT 2925 2930
FT REPEAT 2929 2934
FT REPEAT 2933 2938
FT REPEAT 2937 2942
FT REPEAT 2941 2946
FT REPEAT 2945 2950
FT REPEAT 2949 2954
FT REPEAT 2953 2958
FT REPEAT 2957 2962
FT REPEAT 2961 2966
FT REPEAT 2965 2970
FT REPEAT 2969 2974
FT REPEAT 2973 2978
FT REPEAT 2977 2982
FT REPEAT 2981 2986
FT REPEAT 2985 2990
FT REPEAT 2989 2994
FT REPEAT 2993 2998
FT REPEAT 2997 3002
FT REPEAT 3001 3006
FT REPEAT 3005 3010
FT REPEAT 3009 3014
FT REPEAT 3013 3018
FT REPEAT 3017 3022
FT REPEAT 3021 3026
FT REPEAT 3025 3030
FT REPEAT 3029 3034
FT REPEAT 3033 3038
FT REPEAT 3037 3042
FT REPEAT 3041 3046
FT REPEAT 3045 3050
FT REPEAT 3049 3054
FT REPEAT 3053 3058
FT REPEAT 3057 3062
FT REPEAT 3061 3066
FT REPEAT 3065 3070
FT REPEAT 3069 3074
FT REPEAT 3073 3078
FT REPEAT 3077 3082
FT REPEAT 3081 3086
FT REPEAT 3085 3090
FT REPEAT 3089 3094
FT REPEAT 3093 3098
FT REPEAT 3097 3102
FT REPEAT 3101 3106
FT REPEAT 3105 3110
FT REPEAT 3109 3114
FT REPEAT 3113 3118
FT REPEAT 3117 3122
FT REPEAT 3121 3126
FT REPEAT 3125 3130
FT REPEAT 3129 3134
FT REPEAT 3133 3138
FT REPEAT 3137 3142
FT REPEAT 3141 3146
FT REPEAT 3145 3150
FT REPEAT 3149 3154
FT REPEAT 3153 3158
FT REPEAT 3157 3162
FT REPEAT 3161 3166
FT REPEAT 3165 3170
FT REPEAT 3169 3174
FT REPEAT 3173 3178
FT REPEAT 3177 3182
FT REPEAT 3181 3186
FT REPEAT 3185 3190
FT REPEAT 3189 3194
FT REPEAT 3193 3198
FT REPEAT 3197 3202
FT REPEAT 3201 3206
FT REPEAT 3205 3210
FT REPEAT 3209 3214
FT REPEAT 3213 3218
FT REPEAT 3217 3222
FT REPEAT 3221 3226
FT REPEAT 3225 3230
FT REPEAT 3229 3234
FT REPEAT 3233 3238
FT REPEAT 3237 3242
FT REPEAT 3241 3246
FT REPEAT 3245 3250
FT REPEAT 3249 3254
FT REPEAT 3253 3258
FT REPEAT 3257 3262
FT REPEAT 3261 3266
FT REPEAT 3265 3270
FT REPEAT 3269 3274
FT REPEAT 3273 3278
FT REPEAT 3277 3282
FT REPEAT 3281 3286
FT REPEAT 3285 3290
FT REPEAT 3289 3294
FT REPEAT 3293 3298
FT REPEAT 3297 3302
FT REPEAT 3301 3306
FT REPEAT 3305 3310
FT REPEAT 3309 3314
FT REPEAT 3313 3318
FT REPEAT 3317 3322
FT REPEAT 3321 3326
FT REPEAT 3325 3330
FT REPEAT 3329 3334
FT REPEAT 3333 3338
FT REPEAT 3337 3342
FT REPEAT 3341 3346
FT REPEAT 3345 3350
FT REPEAT 3349 3354
FT REPEAT 3353 3358
FT REPEAT 3357 3362
FT REPEAT 3361 3366
FT REPEAT 3365 3370
FT REPEAT 3369 3374
FT REPEAT 3373 3378
FT REPEAT 3377 3382
FT REPEAT 3381 3386
FT REPEAT 3385 3390
FT REPEAT 3389 3394
FT REPEAT 3393 3398
FT REPEAT 3397 3402
FT REPEAT 3401 3406
FT REPEAT 3405 3410
FT REPEAT 3409 3414
FT REPEAT 3413 3418
FT REPEAT 3417 3422
FT REPEAT 3421 3426
FT REPEAT 3425 3430
FT REPEAT 3429 3434
FT REPEAT 3433 3438
FT REPEAT 3437 3442
FT REPEAT 3441 3446
FT REPEAT 3445 3450
FT REPEAT 3449 3454
FT REPEAT 3453 3458
FT REPEAT 3457 3462
FT REPEAT 3461 3466
FT REPEAT 3465 3470
FT REPEAT 3469 3474
FT REPEAT 3473 3478
FT REPEAT 3477 3482
FT REPEAT 3481 3486
FT REPEAT 3485 3490
FT REPEAT 3489 3494
FT REPEAT 3493 3498
FT REPEAT 3497 3502
FT REPEAT 3501 3506
FT REPEAT 3505 3510
FT REPEAT 3509 3514
FT REPEAT 3513 3518
FT REPEAT 3517 3522
FT REPEAT 3521 3526
FT REPEAT 3525 3530
FT REPEAT 3529 3534
FT REPEAT 3533 3538
FT REPEAT 3537 3542
FT REPEAT 3541 3546
FT REPEAT 3545 3550
FT REPEAT 3549 3554
FT REPEAT 3553 3558
FT REPEAT 3557 3562
FT REPEAT 3561 3566
FT REPEAT 3565 3570
FT REPEAT 3569 3574
FT REPEAT 3573 3578
FT REPEAT 3577 3582
FT REPEAT 3581 3586
FT REPEAT 3585 3590
FT REPEAT 3589 3594
FT REPEAT 3593 3598
FT REPEAT 3597 3602
FT REPEAT 3601 3606
FT REPEAT 3605 3610
FT REPEAT 3609 3614
FT REPEAT 3613 3618
FT REPEAT 3617 3622
FT REPEAT 3621 3626
FT REPEAT 3625 3630
FT REPEAT 3629 3634
FT REPEAT 3633 3638
FT REPEAT 3637 3642
FT REPEAT 3641 3646
FT REPEAT 3645 3650
FT REPEAT 3649 3654
FT REPEAT 3653 3658
FT REPEAT 3657 3662
FT REPEAT 3661 3666
FT REPEAT 3665 3670
FT REPEAT 3669 3674
FT REPEAT 3673 3678
FT REPEAT 3677 3682
FT REPEAT 3681 3686
FT REPEAT 3685 3690
FT REPEAT 3689 3694
FT REPEAT 3693 3698
FT REPEAT 3697 3702
FT REPEAT 3701 3706
FT REPEAT 3705 3710
FT REPEAT 3709 3714
FT REPEAT 3713 3718
FT REPEAT 3717 3722
FT REPEAT 3721 3726
FT REPEAT 3725 3730
FT REPEAT 3729 3734
FT REPEAT 3733 3738
FT REPEAT 3737 3742
FT REPEAT 3741 3746
FT REPEAT 3745 3750
FT REPEAT 3749 3754
FT REPEAT 3753 3758
FT REPEAT 3757 3762
FT REPEAT 3761 3766
FT REPEAT 3765 3770
FT REPEAT 3769 3774
FT REPEAT 3773 3778
FT REPEAT 3777 3782
FT REPEAT 3781 3786
FT REPEAT 3785 3790
FT REPEAT 3789 3794
FT REPEAT 3793 3798
FT REPEAT 3797 3802
FT REPEAT 3801 3806
FT REPEAT 3805 3810
FT REPEAT 3809 3814
FT REPEAT 3813 3818
FT REPEAT 3817 3822
FT REPEAT 3821 3826
FT REPEAT 3825 3830
FT REPEAT 3829 3834
FT REPEAT 3833 3838
FT REPEAT 3837 3842
FT REPEAT 3841 3846
FT REPEAT 3845 3850
FT REPEAT 3849 3854
FT REPEAT 3853 3858
FT REPEAT 3857 3862
FT REPEAT 3861 3866
FT REPEAT 3865 3870
FT REPEAT 3869 3874
FT REPEAT 3873 3878
FT REPEAT 3877 3882
FT REPEAT 3881 3886
FT REPEAT 3885 3890
FT REPEAT 3889 3894
FT REPEAT 3893 3898
FT REPEAT 3897 3902
FT REPEAT 3901 3906
FT REPEAT 3905 3910
FT REPEAT 3909 3914
FT REPEAT 3913 3918
FT REPEAT 3917 3922
FT REPEAT 3921 3926
FT REPEAT 3925 3930
FT REPEAT 3929 3934
FT REPEAT 3933 3938
FT REPEAT 3937 3942
FT REPEAT 3941 3946
FT REPEAT 3945 3950
FT REPEAT 3949 3954
FT REPEAT 3953 3958
FT REPEAT 3957 3962
FT REPEAT 3961 3966
FT REPEAT 3965 3970
FT REPEAT 3969 3974
FT REPEAT 3973 3978
FT REPEAT 3977 3982
FT REPEAT 3981 3986
FT REPEAT 3985 3990
FT REPEAT 3989 3994
FT REPEAT 3993 3998
FT REPEAT 3997 4002
FT REPEAT 4001 4006
FT REPEAT 4005 4010
FT REPEAT 4009 4014
FT REPEAT 4013 4018
FT REPEAT 4017 4022
FT REPEAT 4021 4026
FT REPEAT 4025 4030
FT REPEAT 4029 4034
FT REPEAT 4033 4038
FT REPEAT 4037 4042
FT REPEAT 4041 4046
FT REPEAT 4045 4050
FT REPEAT 4049 4054
FT REPEAT 4053 4058
FT REPEAT 4057 4062
FT REPEAT 4061 4066
FT REPEAT 4065 4070
FT REPEAT 4069 4074
FT REPEAT 4073 4078
FT REPEAT 4077 4082
FT REPEAT 4081 4086
FT REPEAT 4085 4090
FT REPEAT 4089 4094
FT REPEAT 4093 4098
FT REPEAT 4097 4102
FT REPEAT 4101 4106
FT REPEAT 4105 4110
FT REPEAT 4109 4114
FT REPEAT 4113 4118
FT REPEAT 4117 4122
FT REPEAT 4121 4126
FT REPEAT 4125 4130
FT REPEAT 4129 4134
FT REPEAT 4133 4138
FT REPEAT 4137 4142
FT REPEAT 4141 4146
FT REPEAT 4145 4150
FT REPEAT 4149 4154
FT REPEAT 4153 4158
FT REPEAT 4157 4162
FT REPEAT 4161 4166
FT REPEAT 4165 4170
FT REPEAT 4169 4174
FT REPEAT 4173 4178
FT REPEAT 4177 4182
FT REPEAT 4181 4186
FT REPEAT 4185 4190
FT REPEAT 4189 4194
FT REPEAT 4193 4198
FT REPEAT 4197 4202
FT REPEAT 4201 4206
FT REPEAT 4205 4210
FT REPEAT 4209 4214
FT REPEAT 4213 4218
FT REPEAT 4217 4222
FT REPEAT 4221 4226
FT REPEAT 4225 4230
FT REPEAT 4229 4234
FT REPEAT 4233 4238
FT REPEAT 4237 4242
FT REPEAT 4241 4246
FT REPEAT 4245 4250
FT REPEAT 4249 4254
FT REPEAT 4253 4258
FT REPEAT 4257 4262
FT REPEAT 4261 4266
FT REPEAT 4265 4270
FT REPEAT 4269 4274
FT REPEAT 4273 4278
FT REPEAT 4277 4282
FT REPEAT 4281 4286
FT REPEAT 4285 4290
FT REPEAT 4289 4294
FT REPEAT 4293 4298
FT REPEAT 4297 4302
FT REPEAT 4301 4306
FT REPEAT 4305 4310
FT REPEAT 4309 4314
FT REPEAT 4313 4318
FT REPEAT 4317 4322
FT REPEAT 4321 4326
FT REPEAT 4325 4330
FT REPEAT 4329 4334
FT REPEAT 4333 4338
FT REPEAT 4337 4342
FT REPEAT 4341 4346
FT REPEAT 4345 4350
FT REPEAT 4349 4354
FT REPEAT 4353 4358
FT REPEAT 4357 4362
FT REPEAT 4361 4366
FT REPEAT 4365 4370
FT REPEAT 4369 4374
FT REPEAT 4373 4378
FT REPEAT 4377 4382
FT REPEAT 4381 4386
FT REPEAT 4385 4390
FT REPEAT 4389 4394
FT REPEAT 4393 4398
FT REPEAT 4397 4402
FT REPEAT 4401 4406
FT REPEAT 4405 4410
FT REPEAT 4409 4414
FT REPEAT 4413 4418
FT REPEAT 4417 4422
FT REPEAT 4421 4426
FT REPEAT 4425 4430
FT REPEAT 4429 4434
FT REPEAT 4433 4438
FT REPEAT 4437 4442
FT REPEAT 4441 4446
FT REPEAT 4445 4450
FT REPEAT 4449 4454
FT REPEAT 4453 4458
FT REPEAT 4457 4462
FT REPEAT 4461 4466
FT REPEAT 4465 4470
FT REPEAT 4469 4474
FT REPEAT 4473 4478
FT REPEAT 4477 4482
FT REPEAT 4481 4486
FT REPEAT 4485 4490
FT REPEAT 4489 4494
FT REPEAT 4493 4498
FT REPEAT 4497 4502
FT REPEAT 4501 4506
FT REPEAT 4505 4510
FT REPEAT 4509 4514
FT REPEAT 4513 4518
FT REPEAT 4517 4522
FT REPEAT 4521 4526
FT REPEAT 4525 4530
FT REPEAT 4529 4534
FT REPEAT 4533 4538
FT REPEAT 4537 4542
FT REPEAT 4541 4546
FT REPEAT 4545 4550
FT REPEAT 4549 4554
FT REPEAT 4553 4558
FT REPEAT 4557 4562
FT REPEAT 4561 4566
FT REPEAT 4565 4570
FT REPEAT 4569 4574
FT REPEAT 4573 4578
FT REPEAT 4577 4582
FT REPEAT 4581 4586
FT REPEAT 4585 4590
FT REPEAT 4589 4594
FT REPEAT 4593 4598
FT REPEAT 4597 4602
FT REPEAT 4601 4606
FT REPEAT 4605 4610
FT REPEAT 4609 4614
FT REPEAT 4613 4618
FT REPEAT 4617 4622
FT REPEAT 4621 4626
FT REPEAT 4625 4630
FT REPEAT 4629 4634
FT REPEAT 4633 4638
FT REPEAT 4637 4642
FT REPEAT 4641 4646
FT REPEAT 4645 4650
FT REPEAT 4649 4654
FT REPEAT 4653 4658
FT REPEAT 4657 4662
FT REPEAT 4661 4666
FT REPEAT 4665 4670
FT REPEAT 4669 4674
FT REPEAT 4673 4678
FT REPEAT 4677 4682
FT REPEAT 4681 4686
FT REPEAT 4685 4690
FT REPEAT 4689 4694
FT REPEAT 4693 4698
FT REPEAT 4697 4702
FT REPEAT 4701 4706
FT REPEAT 4705 4710
FT REPEAT 4709 4714
FT REPEAT 4713 4718
FT REPEAT 4717 4722
FT REPEAT 4721 4726
FT REPEAT 4725 4730
FT REPEAT 4729 4734
FT REPEAT 4733 4738
FT REPEAT 4737 4742
FT REPEAT 4741 4746
FT REPEAT 4745 4750
FT REPEAT 4749 47
```

```
QY 1750 CAGACGGTCCCTGGTACAACTAT-----CATACGACAGCAAGCGCGTATTGCGAG 1800
Db 693 GlulysThrGlnTyArgSerTyArgGluPheThrHisIleAsnGlyLysAsnLeuThrGlu 712
QY 1801 AAGCAGGAT-----CCTAAGACCGATGAGTTTGCC 1830
Db 713 ThrAspAsnLeuTySerValGluLeuIleGlyThrThrArgAlaAspLysPhe--- 731
QY 1831 GTACAGGAGAAATATCTCTCGACCTCTTC----- 1857
Db 732 PheGlySerLysPheThrAspIlePheHisGlyAlaAspGlyAspLeuIleGluGly 751
QY 1858 CGACACAAAACATAGTCTCCTATCGGACAGCGTCCATCGCTTCCTATCG 1917
Db 752 AsnAspGlyAsnArgLeuTyArgLysPheGly----- 763
QY 1918 AACAAACACATACAGGCGGTAACTCCGCTAGCCCGTATTC-----GATAAGACGGC 1971
Db 764 AsnAspThrLeuSerGlyLysAsnGlyAspAspGlnLeuTyArgLysPheGlyAsnAsp 783
QY 1972 CGTCTGATCGGTCTTCTGCTGATGCGCAACTGGGAGCTATGAGTGGTGACATCGAGTTC 2031
Db 784 LysLeuIleGlyValAlaGlyAsnAsnTyLeuAsnGlyGlyAspGlyAspGluPhe 803
QY 2032 GAA 2034
Db 804 Gln 804

RESULT 14
TEAL_SCHPO
ID TEAL_SCHPO STANDARD; PRT; 1147 AA.
AC P87061;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Tip elongation aberrant protein 1 (Cell polarity protein teal).
GN TEAL OR SPC123.06.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97344085; PubMed=9200612;
RA Mata J., Nurse P.;
RT "teal and the microtubular cytoskeleton are important for generating
RL global spatial order within the fission yeast cell.";
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Spours J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins S., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squires R., Squires S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkhaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
```

```
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: CELL POLARITY PROTEIN. MAY ACT AS AN END MARKER,
CC DIRECTING THE GROWTH MACHINERY TO THE CELL POLES. MAY ALSO
CC INFLUENCE MICROTUBULAR ORGANIZATION, AFFECTING THE MAINTENANCE OF
CC A SINGLE CENTRAL AXIS.
CC -1- SUBCELLULAR LOCATION: PRESENT AT BOTH POLES OF THE CELL THROUGHOUT
CC THE CELL CYCLE WHETHER THEY ARE GROWING OR NOT. LOCATED AT THE
CC ENDS OF MICROTUBULES GROWING TOWARDS THE CELL POLES. AN INTACT
CC MICROTUBULAR SKELETON IS REQUIRED TO MAINTAIN THE LOCATION AT THE
CC CELL TIPS; ON THE OTHER HAND, A NORMALLY ORGANIZED ACTIN
CC CYTOSKELETON IS NOT REQUIRED.
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y12709; CAAT3246.1; -
CC EMBL; AL031579; CAA20875.1; -
CC InterPro: IPR001798: Kelch.
CC Pfam; PF01344; Kelch; 5.
CC Repeat; Microtubules; Coiled coil.
CC REPEAT 94 144 KELCH 1.
CC REPEAT 146 198 KELCH 2.
CC REPEAT 254 303 KELCH 3.
CC REPEAT 305 351 KELCH 4.
CC REPEAT 355 402 KELCH 5.
CC DOMAIN 611 649 COILED COIL (POTENTIAL).
CC DOMAIN 716 838 COILED COIL (POTENTIAL).
CC DOMAIN 879 990 COILED COIL (POTENTIAL).
CC DOMAIN 1084 1105 COILED COIL (POTENTIAL).
CC SEQUENCE 1147 AA; 127436 MW; 7BE55F6C666EF4F8 CRC64;
SQ
Alignment Scores:
Pred. No.: 4.53 Length: 1147
Score: 117.00 Matches: 92
Percent Similarity: 36.99% Conservative: 70
Best Local Similarity: 21.00% Mismatches: 168
Query Match: 3.06% Indels: 109
DB: 16 Gaps: 16
US-10-008-355-1 (1-2139) x TEAL_SCHPO (1-1147)
QY 473 AGGTATCCCAAGAACTGCCCAAAAAGAAATGCAGACGAGCAACCACTCTGCATCGTAG 532
Db 460 ArgAlaSerAsnAspLeuProSerProValValProThrArgSerAsnSerSer----- 477
QY 533 AGCCCTCTTATTCACCAACCAATATTCCTCATCGTCTAGC----- 574
Db 478 SerIleLeuGlnProSerTyArgLeuAsnSerHisSerSerAspArgArgAsnThrAsn 497
QY 575 -----ATGTATTCAAGACAGCTTCCTATGCTAT 601
Db 498 AspAspAspGlnSerSerLeuAsnSerGlnGlnLeuSerAsnGlnAlaLysAlaGlnGly 517
QY 602 TTGCTCTCCAGCTCTG---TAGGTAACTTCGGAGGCGGATACGGACACACTGGATGTCG 658
Db 518 GluValSerProThrLeuSerPheValProSerSerHisSerMetGluGlnGlyAsnGly 537
QY 659 CGCGTACACGCGGAGCTTCAGCGCTATTTCGCGGTATGTCGCGGCGCCGACACCGCGCG 718
Db 538 SerValAlaSerAlaAsnAsnAlaGlnSerGluAla----- 549
QY 719 CCGAATACAGCAAGACAAATAAACCCCTATAAGCCCGTTTACTTCGCTCGCGTATCCATGC 778
```



```
Db 550 -----AlaThrArgSerIleAsnSerIleSerGluValSerGluValArgPheProGlu 567
      |||:::||||| ||||| ::| ::| ::|
Qy 779 AAGGCTACAGGCTGACGACATGCGATGACCATCGTGTTCCTGGGCGCATCGGATCGCT 838
      ::| ::| ::| ::| ::|
Db 568 Gln-----SerValIleThrValAspGluArg 577
      ::| ::| ::| ::|
Qy 839 ACCTCACTTCTGGGTGTGGAAGATGCTATCGAAAGAGAGAACAAATCTCTGATCGAAG 898
      ||||| ||| ::|::| ||| ||| ||||| |||
Db 578 LysSerLeuAspGly-----ArgIleThrSerValThrLeuGluThrLeuValGluLys 595
      ||||| ||| ::|::| ||| ||| ||||| |||
Qy 899 TTCGGCGGTATCAAGCAAGGCATC-----TGAAGGAAGCC-----ATG 936
      ::|::| ||||| ||| ::|::|
Db 596 TyrSerGluLeuSerLysGlnIleValGluTyrPheLysSerLysLeuTyrGluIle 615
      |||||::| ||| ::|::| ::|
Qy 937 AGCGCAGATCAGGCTACCGGTATCAAAATATGCCACAGATATGCTCAGAGTGCTAACTAT 996
      ||| ||| |||::: ||| ::|::|
Db 616 LeuArgAspSerAlaSerLysIleAspSerLeuThrGluLysLeuLysValAlaAsnAla 635
      ::|::| ||||| ||| ::|::|
Qy 997 TGGAAAGATTGATCGGTCGATGAACCGCGGTCTCGCTCTGCTGTGACGTGATAGTGGTAAG 1056
      |||||::| ||| ::|::| ::|
Db 636 GluLysAsnAlaLeuLeuLysGluAlaLeuGluLysValProLeuAlaLysHisAsn 655
      |||||::| ||| ::|::| ::|
Qy 1057 CGTCCGAGGAAGAGCATCGCAGACTGGATCCGTTAAGACGCAAGAGTGCT----- 1110
      ::| ::| |||::| ::|::|
Db 656 LysLeuSerAspGlyThrPheSerThrProAspLysGluAsnValGlnSerThrAsnAsp 675
      ::|::| ||||| ||| ::|::|
Qy 1111 -----GTCATGCGCATGTA 1125
      ||| ::|
Db 676 AlaHisIleMetGlnGluAsnPheSerLeuHisLysAlaLeuGluValMetArgGluThr 695
      |||::| |||::|
Qy 1126 TTGCTCTCTCGAAAGAGCTTATGAAGAGGAGGCAAGCGGAGATGACTTAT 1185
      ||| |||::| |||::|
Db 696 SerSerAspLeuAspLysGlnLysAspAlaThrAlaSerGlnLysGluLeu----- 713
      ||| |||::| |||::|
Qy 1186 TTGACGAGAGAGCTCTCGGTGGTACCAGAGTGTTCTTTTTCACAGTTTGCCAACGCA 1245
      ::|::|::| ::|
Db 714 -----IleValGlnThrSerPheGlnLysGlu 723
      |||::| |||::|
Qy 1246 TTGGCTACAAATCCTGATGCTCATGCTCCCGGTATCCTCAATCGCTTGACGACAGTACAAA 1305
      ||| ::| ||| ||| ||| |||
Db 724 LeuValGluGluArgGluArgHisAsnAlaIleSerLysArgLeuGlnGlu----- 740
      |||::| ||| ||| ||| |||
Qy 1306 GACTACCTCCCTCCCTC-----GACCGTAAGGTGCTGCCCGCATGCTCGATATTGTA 1359
      ::| ||||| |||||::|
Db 741 -----IleGluSerLeuTyrArgAspArgGluLeuLeuValThrAsnLeuGlu----- 756
      |||::| ||||| |||||
Qy 1360 CGCGCGGTATCCCTGCGGACAGCTCCCGCATATATTCAAGATGTATCGACAGAGAA 1419
      |||::| |||
Db 757 -----AspGlnLeuValAspGlnThrValThrIleAsnLysPheAla 770
      |||::| |||
Qy 1420 TTCAAAGCGACACGACGAAGATATGACACTTCGTATTTCGACAGAGTGCTGTTCTTAT 1479
      ||||| ||| ::|::|
Db 771 PheGluArgAspGlnPheArgGluArgSerMetGlyPheGluAsnThrIleLysAspLeu 790
      |||::| |||
Qy 1480 AGCGCAAG-----TTCATGCGCATGCTCAAGTCC 1509
      ::| ||| ||| |||::|
Db 791 ThrArgLysMetGluAlaThrAspMetLeuAsnValSerLeuHisGluSerLeuArgSer 810
      ||| ||| ||| |||
Qy 1510 ATGGACAGAGAAAGTTTGCCAGAGCTATCGAGAAGATCCGGCAGTAGAGCTTTCCAG 1569
      ::| ||| |||
Db 811 ValGlnThrGluAsnSerGluLeuValThrGlu-----MetAlaLeuLeuLys 826
      ||| |||
Qy 1570 AGCGTAATAGTCTGCTCGCGCTATTTCAGGCGCATGCGTGGCAATGCCCTAT 1623
      ::|::| ||||| |||||
Db 827 AlaGluLeuValLysLysGlnAlaIle---IleAspAlaAsnAlaAsnIleTyr 843
      |||::| ||||| |||||
RESULT 15
YK82_YEAST STANDARD; PRT; 1169 AA.
ID YK82_YEAST
AC P36170;
DT 01-JUN-1994 (Rel. 29, Created)
DB 01-JUN-1994 (Rel. 29, Last sequence update)
```

```
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 122.2 kDa protein in SIR1 3'region precursor.
GN YKR102W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Gallon L., Dujon B.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY).
CC -|- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
CC -|- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z28327; CAA82182.1; -.
DR PIR: S38181; S38181.
DR SGD: S0001810; YKR102W.
KW Hypothetical protein; Glycoprotein; Repeat; Cell wall; Signal.
FT SIGNAL 1 26
FT CHAIN 27 1169
FT DOMAIN 274 609
FT CARBOHYD 122 122
FT CARBOHYD 157 157
FT CARBOHYD 279 279
FT CARBOHYD 389 389
FT CARBOHYD 452 452
FT CARBOHYD 515 515
FT CARBOHYD 578 578
FT CARBOHYD 656 656
FT CARBOHYD 879 879
FT CARBOHYD 1092 1092
FT CARBOHYD 1099 1099
SQ SEQUENCE 1169 AA; 122164 MW; 940C6DFB3569C669 CRC64;
Alignment Scores:
Pred. No.: 4.92 Length: 1169
Score: 116.50 Matches: 161
Percent Similarity: 33.56% Conservative: 87
Best Local Similarity: 21.79% Mismatches: 287
Query Match: 3.05% Indels: 204
DB: 1 Gaps: 31
US-10-008-355-1 (1-2139) x YK82_YEAST (1-1169)
Qy 89 TCCTCAACGAACACTCAGGAGAACTCTGGAGTAATCGTGAGCTCGGCTTTACGCTCC 148
      ||||| ::|::| ::|
Db 392 SerSerPheSerSerSerGluValCysThrGluThr----- 407
Qy 149 CGTTGGATTCCCTACAGTTTCGACAGCGCTCCA-----TTGCCAATCCCGTGTTA 202
      ::| ||||| |||
Db 408 -----GluSerThrThrThrThrThrProTyrValThrSerSerSerTrpSer 423
      |||::|
Qy 203 TCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCGAGGCTGTATCTTTACCAACC 262
      ||||| ::|::| |||
Db 424 SerSerGluVal-----CysThrGluCys-----Thr 432
      |||::|
Qy 263 ACCACTCGGGATACGGTGTCTTCCAGAGCCAAAGACGGTGGATCAGACTATCTGCCGC 322
      ||| ||| |||
Db 433 GluThrGluSerThr-SerTyrValThrProTyrValSerSerSerThrAlaAlaAs 452
      ||| |||
Qy 323 ATGTTTCTGTTCTCGCACGATGGGTGAGGACCTCCGATCCGGGCTCTTTCCGTGAAGT 382
      ::| |||
Db 452 nTyr-----ThrSerSerPheSerSerSerSerGluVa 463
      ||| |||||
```





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:21:16 ; Search time 164.5 Seconds  
(without alignments)  
5358.478 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 3820

Sequence: 1 atgcaaatgaataaaag.....aagagctgaagtgtatctaa 2139

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p,model -DEV=xlsp  
-Q/cgn2\_1/USPTO.spool/US1008355/runat\_16052003\_110338\_9352/app\_query.fasta\_1.2311  
-DB=SPREMBL\_21 -Qfmt=fasten -SURFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -START-pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -NAXLEN=2000000000  
-USER=US1008355.ccn\_1.1.238 @runat\_16052003\_110338\_9352 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvivirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	907.5	23.8	716	16 Q9PC94	Q9pc94 xylella fas

2	206	5.4	1349	4	Q8WWMQ4
3	177	4.6	800	3	Q8TFG4
4	157	4.1	957	4	O14651
5	157	4.1	1217	4	Q9UKW9
6	156.5	4.1	1029	4	Q8TDH7
7	156	4.1	1275	4	Q9UQ36
8	156	4.1	1793	4	O15038
9	156	4.1	1791	4	O60382
10	156	4.1	2752	4	Q9UQ35
11	154.5	4.0	1236	3	Q9C105
12	151	4.0	2448	4	Q8WWMQ5
13	149	3.9	528	6	Q29071
14	147	3.8	1037	3	Q8WZX1
15	146.5	3.8	1368	16	Q99ZW2
16	141.5	3.7	1795	5	O76894
17	141	3.7	948	3	O74346
18	139.5	3.7	513	3	O43418
19	139.5	3.7	1802	3	Q04051
20	138	3.5	684	5	Q9VTT7
c 21	137	3.6	3570	4	Q99552
c 22	136	3.4	744	3	Q8TFG9
23	135.5	3.5	589	4	O43419
24	135.5	3.5	2296	4	Q9UHA8
c 25	135	3.4	521	17	Q96Z01
26	134	3.5	477	4	Q14887
27	133.5	3.5	1104	2	Q06812
28	131.5	3.4	901	4	Q9H195
29	131.5	3.4	991	17	O8ZVU0
30	131	3.4	648	4	O14760
31	131	3.4	744	3	Q8TFG9
32	130	3.4	1175	16	Q8RHT3
33	129.5	3.4	1182	5	Q9U7F0
34	127	3.3	2232	5	P91365
35	126.5	3.3	666	17	O27798
36	126.5	3.3	3178	5	Q969D4
37	126	3.3	953	2	Q48502
38	125	3.3	1440	5	Q9TYU6
39	125	3.3	2055	5	Q8T5C7
40	124.5	3.3	735	16	O8U8L4
41	124.5	3.3	799	5	O18119
42	124	3.2	1325	5	Q9BKV7
43	124	3.2	1808	10	Q9LQJ8
44	123.5	3.2	713	2	O85051
45	123	3.2	1289	16	Q9X087

ALIGNMENTS

RESULT 1

Q9PC94	ID	Q9PC94	PRELIMINARY;	PRT;	716 AA.
AC	Q9PC94;				
DT	01-OCT-2000	(TRENBLrel. 15, Created)			
DT	01-OCT-2000	(TRENBLrel. 15, Last sequence update)			
DT	01-JUN-2002	(TRENBLrel. 21, Last annotation update)			
DE	Hypothetical protein Xf1887.				
GN	Xf1887.				
OS	Xylella fastidiosa				
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;				
OC	Xylella.				
OX	NCBI_TaxID=2371;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=9A5C;				
RX	MEDLINE=20365717; PubMed=10910347;				
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,				
RA	Alvaranga R., Alves L.M.C., Araya J.E., Balá G.S., Baptista C.S.,				
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,				
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,				
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,				
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,				
RA	Faccinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,				
RA	Fraga J.S., Franca S.C., Frohme M., Furlan L.R.,				

RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Hohenisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
RA Vallada H., Van Slys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RL "The genome sequence of the plant pathogen *Xylella fastidiosa*;"  
DR Nature 406:151-159(2000).  
DR EMBL: AE004008; AAF84693.1; ..  
DR MEROPS: S46.001; ..  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 716 AA; 79375 MW; EAF086E2315BBDPC CRC64;

## Alignment Scores:

Pred. No.: 1,02e-54 Length: 716  
Score: 907.50 Matches: 239  
Percent Similarity: 49.21% Conservative: 135  
Best Local Similarity: 31.45% Mismatches: 291  
Query Match: 23.76% Indels: 95  
DB: 16 Gaps: 21

US-10-008-355-1 (1-2139) x Q9PC94 (1-716)

Qy 1 ATGCAATGAATTAAGAATTAATCTTCGAGAGCCCTCGTGTGGTGTTCAGGG 60  
Db 1 MetArgPheAsnLeuSerLeuSerValLeuAlaThrLeuIleThrValAspSer--- 19  
Qy 61 GTAGCCAAAGCCGACAGGAGTGTGCTCTCAACGAACTCAATCAGGAGAACTCGAT 120  
Db 20 ---ThrHisAlaGlyGlyMetTrpValProGlnGlnLeu---ProGluIleAlaGly 37  
Qy 121 CGAATGCGTACGCTTACGCTCCGCTGCGTGTGATTCACAGTTTCAGAGTTTCAG 180  
Db 38 ProLeuLysGlnAlaGlyLeuGlnLeuSerProGluGlnLeuSerAsnLeuThrGlyAsp 57  
Qy 181 TCCATTGCCAATGCGGTGTTATCTTCGGTGGCGGATGTACCGGTATCAGAGTCCGAT 240  
Db 58 ProMetGlyAlaValValSerLeu-----GlyAsnCysThrAlaSerLeuValSerPro 75  
Qy 241 CAGGCGCTGATCTTTACCAACCCACCTCGCGATACGCTGCTATCCAGAGCAAGCAGC 300  
Db 76 GluGlyLeuValIleThrAsnHisCysAlaThrGlyAlaIleGlnLeuAsnSerThr 95  
Qy 301 GTGATCACGACTATCTCGCGATGGTTCGTTCTCGCAGGATGGGTGAGGAGCTTCGG 360  
Db 96 ProLysLysAsnLeuIleLysGluGlyPheAsnAlaLeuThrGlnAlaAspGluValSer 115  
Qy 361 ATT---CGGGTCTTTCGTAAGTATCTCGCGAAGATCGTGAAGGTAAACGACAGGTA 417  
Db 116 AlaGlyProAsnAlaArgIleThrValLeuGluGlnIleThrAspValThrAlaGlnAla 135  
Qy 418 GAAGGACAGCTCAAGGTTATCACTGACGAGATGCGAGGCTCTGCCCAAGCTCAGAGGTA 477  
Db 136 LysAlaAlaLeuAlaAlaGlyAsnAspProPheLysArgThrThrAlaLeuGluThr 155  
Qy 478 TGC-----CAAGAACTGGCCAAAAGAAAATGTCAGACAGAGAACCAACTCTGCATCGTA 531  
Db 156 PheSerLysGlnGluIleAlaLysCysGlu---GluGlnGlyTyrArgCysGlnPhe 174  
Qy 532 GAGCCTTCTATCCACAGCAATACTTCTCTACGATCTACGATGATTCAACGAGCTT 591  
Db 501 -----LeuLysTrpPheAsnAlaAspArgAlaAlaPheGluSerSerGln 515

Db 175 PheSerPheAlaGlyGlyAsnThrTyrArgValPheLysAsnLeuGluIleLysAspVal 194  
Qy 592 CGTATGTTATTTGCTCTCCAGCTCTAGTAAAGTTCGAGCGCATACGACGAACCTGG 651  
Db 195 ArgLeuValTyrAlaProGlnGlySerValGlyLysPheGlyGlyAspValAspAsnTrp 214  
Qy 652 ATGTGGCGCGCTCACACGGCGACTTCAGCGTATTCCGCGCTGTATCCGCGCGCGCAAC 711  
Db 215 MetTrpProArgHisThrGlyAspPheSerPheTyrArgAlaTyrValGlyLysAspGly 234  
Qy 712 CGGCGCGCGGANTACACAGGAGCAATAACCCCTATTAAGCCCGTTTACTTCGCTCCGCTA 771  
Db 235 LysProAlaSerPheSerLysGluAsnIleProTyrArgProLysHisThrLeuLysPhe 254  
Qy 772 TCCATGCAAGGCTCAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGTACG 831  
Db 255 SerAspGlnProLeuGlyAspGlyAspPheValMetValAlaGlyTyrProGlyArgThr 274  
Qy 832 GATCGCTACCTTCATCTTGTGGGTGTGGAGATCGTATCGAAAACGAGAACAACTCCTCGT 891  
Db 275 AsnArgTyrAlaLeuVal-----AlaGluPheGluAsnThrAlaHis 288  
Qy 892 ATCGAAGTTCGCGGTATCAAGCAA-----GCCATCTGGAAGGAAGCC 933  
Db 289 TrpThrTyrProValIleGlyGlnHisPheLysAsnLeuIleAlaLeuIleGluAlaAla 308  
Qy 934 ATGAGCGCAGATCAGGCTACCCGATCAAAATATCCAGCAAGTATGCTCAGAGTGTAAAC 993  
Db 309 SerLysGlnAsnProAspIleGlnValLysTyrAlaSerThrLeuAlaGlyLeuAsnAsn 328  
Qy 994 TATTGGAAGAAATTCGATCGGTATGAACCGCGTCTCGTCTTTCAGCTGATAGTCTCGT 1053  
Db 329 ThrSerLysAsnPheAspGlyGlnLeuAspGlyPheArgArgIleAsnAlaIleGlyGln 348  
Qy 1054 AAGCTCCGAGGAAGAGCAATCCGACACTGGATCGCTAGACGGC-----1101  
Db 349 LysGlnSerGluGluThrAlaValLeuAlaTrpLeuLysGlnGlnGlyIleArgGlyHis 368  
Qy 1102 AAGAGTGTCTATGCGGATGTTGCTCTCTCGAAAAGGCTTATTAAGGAAGGAGCC 1161  
Db 369 GluAlaLeuAlaHisGlnThrLeuValAspLeuThrGluGlnTyr-----384  
Qy 1162 AAGCCCAACCTGAGTACCTTATTGAGCGAGAGCGCTTTCGCTGGTGTACGAGTGTGT 1221  
Db 385 LysAlaAsnGlnAspArgAspPheValLeuGlyGln---PheAsnGlySerGlyValIle 403  
Qy 1222 CGTTTTCACAGTTTGGCAACCGCATTCGCT-----ACAAATCTCGTGTCTCAT--- 1269  
Db 404 GlyValAlaValAsnLeuTyrArgLeuAlaIleGluArgThrLysSerAspAlaGlnArg 423  
Qy 1270 -----GCCGTATCTCAATCTGCTTAC 1293  
Db 424 GluAlaGlyTyrGlnGluArgAspLeuProThrIleGluGlyAsnLeuLysGlnMetGlu 443  
Qy 1294 GACAAGTACAAGACTACCTCCCTCGCTCGACCGTAAGGTG-----1335  
Db 444 -----ArgArgTyrLeuProGluMetAspArgGlnMetGlnTyrTrpLeuThr 460  
Qy 1336 -----CTGCGCGCATCTCGATATTTCAGCCCGCGCTATCCCTGCC-----1377  
Db 461 GluTyrAsnLysLeuPro-----ValLysGlnArgValAlaAlaIleAsp 475  
Qy 1378 -----GACAAGCTCCCGCATATTTCAAGAATGTAATTCGACAAAGAAATTTCAA 1425  
Db 476 ValTrpLeuGlyAspGlyIleProAlaThrLeuLysArgLeuGlyAspThrLysLeuSer 495  
Qy 1426 GGCACACGAGAAAGATATGACAGACTTCGTATTTCGACAAAGAGTGTGTTCTTATAGCGAC 1485  
Db 496 SerSerGluGluArg-----500  
Qy 1486 AAGTTTCCATGCCATGCTCAAGTCCATCGACAGGAAAAGTTTGCAGAGGCTATCGCAGAA 1545  
Db 501 -----LeuLysTrpPheAsnAlaAspArgAlaAlaPheGluSerSerGln 515

```
QY 1546 GATCCGGCAGTAGACGTTTCCAGAGCGTAATAGCTGCTGCTCGCGCTATTTCAGCGCGAT 1605
|||||
Db 516 AsproAlaIleargTyAlaValAlaIleMetProAlaLeuLeuGluLeuArgGln 535
|||||
QY 1606 CGATGCCAATGCTATGCGATGAGAGGCGAAGCGTCTTTTCTTGGCGGTTTCGGT 1665
|||||
Db 536 AsnLysIleargThrGlyGluLeuLeuLysAlaArgProIleTyLeuGlnAlaLeuAla 555
|||||
QY 1666 GAGATGTACCCC-----GGAGGTGCTCTCCGAGCGCATGCCAAGTTCACCATCGCGTATG 1719
|||||
Db 556 AspTyrAsnLysSerHisGlyLysPheValTyProAspAlaAsnSerSerLeuArgIle 575
|||||
QY 1720 AGCTACGGCTCCATCAAGGATATGAACCGAGCGGTCGCTGGTACAACTATCATACG 1779
|||||
Db 576 ThrPheGlyHisValLysGlyTySerProLysAspGlyValGluTyThrProPheThr 595
|||||
QY 1780 ACAGGCAAGGCGTATTTGGAGACGAGCATCTTAAGACGATGATGTTGCGGTACAGGAG 1839
|||||
Db 596 ThrLeuGlnGlyValMetAlaLys---AsnThrGlyValGluProPheAspSerProLys 614
|||||
QY 1840 AATATCTCGACCTCTCCGACCAAACTATGTCGCTATGCGGAGAAC-----GGT 1893
|||||
Db 615 SerLeuIleAsnAlaIleLysAlaLysSerTyAlaAsnLeuAlaAspGlnArgIleGly 634
|||||
QY 1894 CAGCTCATCGCTTCTTCCATCAAGACGACATCACGGGCGGTAACCTCCGGTAGCCCC 1953
|||||
Db 635 ThrValProValAsnPheLeuSerAspLeuAspIleThrGlyGlyAsnSerGlySerPro 654
|||||
QY 1954 GTATTCATGAAGACGGCGCTGATCGGCTGCTGCTTTCGATGCGCACTGGGAAGCTATG 2013
|||||
Db 655 ValLeuAspAlaHisGlyLysLeuValGlyLeuAlaPheAspGlyAsnTrpGluSerVal 674
|||||
QY 2014 AGTGGTCACATGAGTTGCAACCGATCTCGAGCGCAACATCAGCGTGACATCCGCTAC 2073
|||||
Db 675 SerSerAsnTrpValPheAspProValMetThrArgThrIleAlaValAspSerArgTyr 694
|||||
QY 2074 GTTCTCTCATGATGATAAATGGGTGAGTGGCGGCTGCTCATCCAAAGAGCTGAAGTTG 2133
|||||
Db 695 ValGlnTrpIleMetThrGluValAlaProAlaProHisLeuLeuLysGluLeuAsnLeu 714
|||||
RESULT 2
Q8WQ4 PRELIMINARY; PRT; 1349 AA.
AC Q8WQ4;
AT Q8WQ4;
CT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Mucin 5 (Fragment).
GN MUC5AC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21426417; PubMed=11535137;
RA Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
RT repetitive region"; 772(2001).
RL Biochem. J. 358:763-772(2001).
DR EMBL; AJ298318; CAC83675.1; -.
FT NON_TER 1
FT NON_TER 1349 1349
SQ SEQUENCE 1349 AA; 135600 MW; 4DC3C1544F1E5EBA CRC64;
```

## Alignment Scores:

```
Pred. No.: 1.21e-05 Length: 1349
Score: 206.00 Matches: 168
Percent Similarity: 32.71% Conservative: 78
Best Local Similarity: 22.34% Mismatches: 281
Query Match: 5.39% Indels: 225
DB: 4 Gaps: 30
```

```
US-10-008-355-1 (1-2139) x Q8WQ4 (1-1349)
QY 41 TCGTCTTGGGTGCTTTCAGGGGTACCCA-----AAGCCGACAAAG 79
|||||
Db 593 CysGluThrProLysGlyCysProValThrSerThrSerValThrAlaProSerPro 612
|||||
QY 80 GCATGTGGC-----TCCTCAACGAACCAATCAGCAGCAATCTGGATC 121
|||||
Db 613 LeuValGlyGluProProAlaGlnThrGlnSerThrSer-----TrpGln 628
|||||
QY 122 GAATGCTGAGTCGGGTTTACGCTCCCGTTGGATTGCTCTACAGTTTCGACAAGC--- 178
|||||
Db 629 LysSerArgThrThrThrLeuValThrSerSerIleThrSerThrGlnThrSerThr 648
|||||
QY 179 -----CGTCCATTGCCATGCGGTGTTATCTTCGGTGGCG 214
|||||
Db 649 ThrSerAlaProThrThrSerThrProAlaSerIleProSerThrThrSerAlaPro 668
|||||
QY 215 GATGATCCGGTATCACAGTGTCCGATCAGGGCCTGATCTTTACCAAGCACCCACTCGCGAT 274
|||||
Db 669 ThrThrSerThrThrSerAlaPro-----ThrThrSerThrThrSerAlaPro 684
|||||
QY 275 ACGGTGCTATCCAGAGCCAAAGCAGCGTGGATCAGCACTATCTCGCGGATGGTTTGGTT 334
|||||
Db 685 ThrThr---SerThrThrSerThrProGlnThrThrSerSerAlaProThrSerSer 703
|||||
QY 335 CTCGACGATGGGTGAGGAGCTTCGGATTCGGGTCTTCGCTGAAGTATCTCGCGCAAGA 394
|||||
Db 703 ----- 703
QY 395 TCGTGAAGTAACGCAAGGTAGAACGACAGCTCAAGGGTATCCTCAGCAGATGGAGC 454
|||||
Db 704 -----ThrThrSerAlaProThrThrSerThrIleSerAlaProThrThrSer 719
|||||
QY 455 GTCGCGCAAGCTCAGAGGTATGCCAAGAACTGGCCAAAGAAAAAGAAATCGACAGCAGA 514
|||||
Db 720 ThrIleSerAlaProThrThrSerThrSerAlaProThrAlaSerThrThrSerAla 739
|||||
QY 515 ACCAATCTCGATCGTAGACCTTCTTATCCAAACAGAACTCTCTCATCTCTCAGC 574
|||||
Db 740 ProThrSerThrSer-----SerAlaProThrAsnThrThrSerAlaProThr 756
|||||
QY 575 ATGTATTCAAGGAGCTTCGTATGTTGCTCTCCCA---GCTCTGTAGTAGTTCG 631
|||||
Db 757 ThrSerThrThrSer-----AlaProIleThrSerThrIleSerAla 770
|||||
QY 632 GAGCGATACGGACAACCTGATGTCGCGCGCTACACGGGCGACTTCAGCGTATTCGCG 691
|||||
Db 771 ProThrThrSerThrSerThrProGlnThrSerThrIleSerSerProThrThrSer 790
|||||
QY 692 TGTATGCGGTGCGG-----ACAACGGCGCGCGAATACAGCAAGGACAAATTAAC 742
|||||
Db 791 ThrThrProThrProGlnThrSerThrSerSerProThrThrSerThrSerAla 810
|||||
QY 743 CCTATAAGCCCGTTTACTTTCGCTGCTGCTATCCATGCAAGGCTACAGGCTGACGACTATG 802
|||||
Db 811 ProThrThrSerThrThrSerAlaPro-----ThrThrSerThrThrSerThr 826
|||||
QY 803 CCATGACCATCGGTTTCCGGGCGAGTACGGATCGCTACCTCCTCTTCTGGGGTGTGAAG 862
|||||
Db 827 ProGlnThrSerIleSerSerAlaProThrSerSerThrThrSerAlaProThrAlaSer 846
|||||
QY 863 ATCGTATCGAAAGGAGAACAACTCTCGTATCGAAGTTCGGCGGTATCAAGCAAGCATCT 922
|||||
Db 847 ThrIleSerAlaProThrThrSerThrThrPheIleThrThrSer----- 863
|||||
QY 923 GGAAGGAGGCGCATGAGCGCAGATCAGGCTACCGGTATCAAAATATGCCCAAGTATGCTC 982
|||||
Db 863 ----- 863
QY 983 AGACTGCTAACTATTGGAAGAAATTCGATCGGTATGAACCGCGGTCTCGCTCTGTCG 1042
|||||
```

```
Db 863 ----- 863
Qy 1043 TGATAGTCGTAAGCGTCCGAGAAAGAGCATTCGCAGACTGATCCGTAAGAACGCCA 1102
Db 864 -----ThrSerProThr-----SerThrSerSerThrProGlnThr 877
Qy 1103 AGAGTGTGCTATGGGAGTATGCTCTCTCGAAAAAGGCTTATAAGGAAGAGCCCA 1162
Db 878 SerLysThrSerAlaAlaThrSerSerThrThrSerGlySerGlyThrThrProSerPro 897
Qy 1163 AGGCCAACCGTGCATGACATATTATTGAGCGAGAGCGCTCTTCGGTGGTACCGAGGTGCTC 1222
Db 898 ValProThrThr----- 901
Qy 1223 GTTTTGCACAGTTTGCACAGCATTTGGCTACAAATCTGTATGCTCATCGCGGTATCCTCA 1282
Db 902 -----SerThrAlaSerValSerLysThrSerThr 911
Qy 1283 AATCGCTTGAGCAAGTACAAAGACTACCTCCCTCGCTCGAGCGTAAGTGCTGCCCG 1342
Db 912 SerHisValSerValSerLysThrThrHisSerGlnProValThrArgAspCysHisPro 931
Qy 1343 CCATGCTCG---ATATTGTACCGCGGCTATCCCTCGCGACAACTCCCG----- 1390
Db 932 ArgCysThrTrpThrLysTrpPheAspValAspPheProSerProGlyProHisGlyGly 951
Qy 1391 -----ATATATTCAAGATGTAATCGA-----CAAGAAATTCAAAGCGCACACGA 1435
Db 952 AsplysGluThrTyraAsnAnile-IleArgSerGlyGluLysIleCysArgArgProGln 971
Qy 1436 AGAAGTATGCAACTTCGTATTCGACAAAGAGTGTGGTTCCTTATPAGCGACAAAGTTCCTATG 1495
Db 971 uGluIleThrArgLeuGlnCysArgAlaGlu-----CysSerArgGluGluGly-----Se 982
Qy 1496 CCATGCTCAAGTCCATGCACAA-----GGAAAGATTGCAAGGCTATCGAGAAAGATC 1549
Db 982 rHisProGluValSerIleGluHisLeuGlyGlnValGln----- 996
Qy 1550 CGGCAGTAGAGCTTTCCAAAGAGCGTAATAGCTGCTCGCGCTATTTCAGGCCCGATCGCA 1609
Db 997 -----CysSerArgGluGluGly----- 1002
Qy 1610 TGGCCAATGCTATGCCATTTAGAGGCGCAAGCGCTCTTTTTCGCGGTTTTCGCTGAGA 1669
Db 1003 -----LeuValCysArgAsnGlnAspGlnGlnGly--ProPheLysMetCysLeuAsn 1019
Qy 1670 TGTACCCCGGAGCTGCTCCGAGCGCATGCCA-----ACTTCACCATGC 1714
Db 1020 TyrGluValArgValLeuCysCysGluThrProLysGlyCysProValThrSerThrPro 1039
Qy 1715 GTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGCGGTGCTGTACAACTATC 1774
Db 1040 Val-----ThrAlaProSerThrProSerGlyArgAlaThrSerPro----- 1053
Qy 1775 ATACGACAGCAAGGGGTATTGG---AGAAGCAGGATCTTAAGAGCGATGAGTTTCCGG 1831
Db 1054 ---ThrGlnSerThrSerSerTrpGlnLysSerArgThrThrThrLeuValThr----- 1070
Qy 1832 TACAGGAGATATCTCGACTCTCTCCGCAACCAAACTATGGTCGCTATCCCGAGAACG 1891
Db 1071 -----ThrSerThrThrSerThrProGlnThrSerThrThrSerAlaProThr 1086
Qy 1892 GTCAGCTCCATATCGCTTTCTTCTCGAACACAGCATCAGCGGGCGGTAACTCCCGGTA--- 1948
Db 1087 ThrSerThrIleProAlaSerThrProSerThrThrSerAlaProThrThrSerThrThr 1106
Qy 1949 ---CCCGGATTCGATAAGAACGGCC-----GTC 1975
Db 1107 SerAlaProThrThrThrSerThrSerAlaProThrHisArgThrThrSerGlyProThr 1126
Qy 1976 TGATCGGTCTTCTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCCGAAC 2035
Db 1127 ThrSerThrThrLeuAlaProThrThrSerThrThrSerAlaProThrThrSerThrAsn 1146
```

```
Qy 2036 -----CCGATCTGCAGCGCACATCAATCAGCTGGACATCCGCTAGTCTTCTTCATGATTG 2089
Db 1147 SerAlaProThrThrSerThrIleSerAlaSerThrThrSerThrIleSerAlaProThr 1166
Qy 2090 ACAAAATGGGTGCTAGTGCCTCCCGCTCATCC 2119
Db 1167 ThrSerThrIleSerSerProThrSerSer 1176
RESULT 3
O8TFG4 PRELIMINARY; PRT; 800 AA.
ID Q8TFG4 AC Q8TFG4;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Putative glucoamylase, similar to S. cerevisiae STA1, similar to S.
DE pombe SPBPJ4664.02.
GN SPAP18E9.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA O'Neill S., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL691402; CAD27470.1; -;
SQ SEQUENCE 800 AA; 79351 MW; F65488E43FE7089A CRC64;
Alignment Scores:
Pred. No.: 0.000984 Length: 800
Score: 177.00 Matches: 185
Percent Similarity: 34.78% Conservative: 111
Best Local Similarity: 21.74% Mismatches: 315
Query Match: 4.63% Indels: 241
DB: 3 Gaps: 35
US-10-008-355-1 (1-2139) x O8TFG4 (1-800)
```

```
Qy 34 GCAGCCCTGCTGTTGGTGCTTTCAGGGGTAGCCAAAGCCGACAAAGCATGTGGC----- 88
Db 7 SerThrLeuLeuLeuLeuAlaGlnHisLeuAlaProAla-AlaAlaSerCysGlyLeuAl 26
Qy 89 -----TCCTCACGAACCTCAATCAGGAGATCTGGATCGAATCCGTCGAGCTC 135
Db 26 aAsnGluAlaSerLeuSerAlaAsnGluValGlnAspIleTyrHisAspPheValAlaTh 46
Qy 136 GGCTTTA-----CGCTCCCGTTGGATTCCGCTACAGTTTCGAC 174
Db 46 rSerIleAsnTyraAlaAspLeuLeuLysArgAsnGluAspLeuAsnAlaSerLeuSerTh 66
Qy 175 AAGCCGTCCATTCGAATGCCGTGGTTATCTTCGGTGGCGGATGTACCGGTATCAGAGT 234
Db 66 rGlySerProValGluIleThrSerThrSerCysThrThrAspThrSerAlaSer---Th 85
Qy 235 TCCGATCAGGCGCTGACT----- 253
Db 85 rProIleIleThrGluSerThrSerSerThrSerSerAlaSerThrThrGlySerSerSe 105
Qy 254 -----TTACCAACACC-----ACTGCGGATACGGTGCTATCCAGAGCAA 294
Db 105 rSerProLeuProSerThrSerThrSerCysThrThrSerThrSerIle-Pro---Pro 124
Qy 295 AGCAGGTGATCAGACTATCTCCGCGATGGTTTCGTTCTCCACGATGGGTGAGGAG 354
Db 124 hrGlyGlySerSerSerLeuSerThrProIleThrProThrValProProThrSerThrS 144
Qy 355 CTTCGATTCGGGCTCTTCGTCGAAGTATCTCGCAAGATCGGTGAAGGTAACCGACAAG 414
Db 144 erSerThrIleProIleProThrProThrSerSerSerThrSerSerThrAspThrAsnSerAnp 164
```



```
QY 415 GTAGAGGACAGCTCAAGGGTATCACTGACGAGATGGAGGCTCTGCGCAAGAGCTCAGGAG 474
Db 164 roLeuProThr-----SerThrSerCysThrThrSerThrs 177
QY 475 GTATGCCAAGAACTGGCCA-AAAAGAAATGACAGAG-----A 512
Db 177 erIleProThrGlyGlySerSerLeuSerThrProIleThrProThrValProp 197
QY 513 GAACCAACTCTGCATCTAGAGCTTTCTATTCCAAACAGAACTCTCTCATCGCTA 572
Db 197 roThrSerThrSerThrSerIleProIleProProThrSerThrSerThrAspT 217
QY 573 CGATGTATTCAAGGACGCTCGTATGTTGCTCCCTCCAGCTCTGTAGTAAGT--- 628
Db 217 hrAsnSerSerPro-----LeuProThrThrSerThrSerCysT 230
QY 629 -----TCGGAGGCGATAGGCACAACTCGATGTCGCCCGGTACACAGG----- 670
Db 230 hrThrSerThrSerIleProThrGlyGlySerSerLeuSerThrProIleThrProt 250
QY 671 -----GGACTTACAGCGTATCCCGGTGTATGCGCGTGCCG----- 706
Db 250 hrValProProThrSerThrSerThrSerIleProIleProProThrSerThrs 270
QY 707 -----TTTACTTCGCTCGCTATCCATGCAAGGCTACAGGCTGACG-----ACTA 800
Db 270 erThrAspThrAsnSerSerProLeuProThrThrSerThrSerCysThrThrSerThrs 290
QY 714 GCGCGCGAATACAGCAAGACATAAACCCGTATAAGCCG----- 754
Db 290 erIleProProThrGlyAsnSerThrThrProValThrProThrValProProThrSert 310
QY 755 -----TTTACTTCGCTCGCTATCCATGCAAGGCTACAGGCTGACG-----ACTA 800
Db 310 hrSerSerThrThrProProProProAlaSerThrSerSerThrGlyThrSerSers 330
QY 801 TGCCATGACCATCGGTTTCCGGCGAGTACGATCGCTACTCTCTTGGGTGGGA 860
Db 330 erProLeuProSerThrSerThrSerCysThrThrSerThrSerIleProProThrGlyA 350
QY 861 AGATCGTATCGAAGACAGACATCTCGTATCGAAGTTCGCGGTATCAAGCAAGGCAT 920
Db 350 snSerThrThrProVal---ThrProThrValProProThrThrSerSerThrSert 369
QY 921 CTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCCGTATCAAAATGCGCAGCAAGTATGC 980
Db 369 hrProProProProAlaSerThrSerThrGlyThrSerSerSerProLeu----- 386
QY 981 TCAGAGTGCTAACTATTGGAAGAATTGATCGGTATGAACCGCGGTCTCGTCTTTGA 1040
Db 387 -----LeuSerThrSerThrSerCysThrThrs 396
QY 1041 CGTATAGTGTAAAGCGTCCGAGGAAGAGCAATTCGACAGCTGGATCCGTAGAACGG 1100
Db 396 er-----ThrSileProProThrGlyAsnSerThrThrProValThrProThrV 413
QY 1101 CAAGAGTGCTGTATGCGG-----ATGATTGTCTCTCTCGAAAAGGC 1145
Db 413 alProProThrSerSerThrProLeuThrThrThrAsnCysThrThrSerThrs 433
QY 1146 TT-----ATAAGGAAGGAGCAAGCCCAACCGTGAGATGACTTATT 1187
Db 433 alProTyThrSerThrProValThrSerThrProLeuAlaThrThrAsn----- 449
QY 1188 GAGCGAGAGCGTCTTCGGTGTACCGAGGTGTCGTTTCCACAGTTCGCCAAGCAT 1247
Db 450 --CysThrThrSerThrValPro----- 457
QY 1248 GGCTACAAATCCTGATGCTCATGCGGTATCTCAAAATCGCTTCGACGACAGTACAAAGA 1307
Db 458 -----TyThrSerThrProValThrSerThrProLeuThrThrThrAsnCyst 474
```

```
QY 1308 CTACCTCC-----CCTCGCTCGACCG 1328
Db 474 hrThrSerThrSerIleProTyThrSerThrProValThrSerThrProLeuThrThrT 494
QY 1329 TAAGTGCTGCTGCCGCCCATGCTCGATATTGTACGCCGGGTATCCCTGCCGCAAGCTCCC 1388
Db 494 hrAsnCysThrThrSerThrSerValProTy-----ThrSerThrProValThrSers 512
QY 1389 CGATATATTCAAGAATGTAATCGACAAAGAAATTCAAAGGCGCACAGGAAGTAGTCAGA 1448
Db 512 erAsnTyThrIleSerSerSerThr-----ProValThrSerThrProValThrT 529
QY 1449 CTTCGTATTTCGACAAGAGTGTGGTTCCTTATPAGCACAAGTTCATGCCATGCTCAAGTC 1508
Db 529 hrThrAsnCysThrThrSerThrSerValLeuTyThrSerThrProValThrSerThrP 549
QY 1509 CA-----TGGACAAGGAAAGTTTCCCAAGGCTATCGAAGAAGATCC 1550
Db 549 roLeuAlaThrThrAsnCysThrThrSerThrSerValProTyThrSerThrProValT 569
QY 1551 GGCAGTAGAGCTTTCCCAAGAGCGTAATAGTCTGCTCGCGCTATTTCAGGCCGATGCGAT 1610
Db 569 hrSerSerAsnTyThrIleSer----- 576
QY 1611 GGCAATGCTATGCCATTCGAAAGGCAAGCGTCTTTCTTTCCCGGTT---TGCSTGA 1667
Db 577 -----SerSerThrProValThrSerThrProValThrThrThrA 590
QY 1668 GATGTACC-----CCGAGCGTCTCTGCCGAGCGATGCCA---ACTTCACCA----- 1711
Db 590 snCysThrThrSerThrSerValLeuTyThrSerThrProIleThrSerProAsnSert 610
QY 1712 -----TGCGTATGAGCTACGGCTC--- 1730
Db 610 hrSerSerSerSerThrGlnValSerThrPasnSerThrThrProIleThrGlyThrSert 630
QY 1731 CATCAAGGATATGAACCGCAGGACGGTCTGTGTACAACTATCATACGACAGCAAGGG 1790
Db 630 hrSerLysValThrSerSerThrSileProLeuThrSerThrAsnArgThrSerThrT 650
QY 1791 CGTATTGGAGAAGCAGGATCCTAAGACGATGAGTTTGCCG----- 1831
Db 650 hrPheThrSerSerThrSerIleSerThrSerSerSerSerThrAlaThrSerThrs 670
QY 1832 -----TACAGGAGAATA-----TCCTCGACCTCTT 1856
Db 670 erPheAlaSerGluSerSerSerPheTyThrSerAsnValThrThrSerSerThrVals 690
QY 1857 CCGCACCAAAACTATGTCGCTATGCCGAGACGGTCACTCATATCGCTTTCTTATC 1916
Db 690 erThrProProThrThrSerPhePro-----SerThrPheThrThrSerPheI 707
QY 1917 GAACAACGACATCAGCGCGGTAACTCCGGTAGCCCGCTATTCGATTAAGACGCGCTCT 1976
Db 707 leThrSerSerSerLeuSerSerIleProAsnAsnSerThrGluValLysThrAla---- 725
QY 1977 GATCGGTCTTCTTCATGCGAACTGGGAAGCTATGATGGTGCATCGAGTTCGAACC 2036
Db 726 -----SerThrSerSerGlyThrGlu---IleLysThrAlaSerThrSers 740
QY 2037 CGATCTCGACGCGCAATCATACGCTGGACATCCGCTACGCTTCTTTCATGATGCAAAATG 2096
Db 740 erGlySerSerSerSerSerTyThrProAlaSerSerThrSerThrThrThrSers 760
QY 2097 GGGTCAGTCCCGCTCTCATCCCAAGAC 2125
Db 760 erValSerSerArgGlnSerSerSer 769
RESULT 4
O14651 PRELIMINARY; PRT: 957 AA.
AC O14651; O14650;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
```

DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	Mucin (intestinal mucin) (Fragment).	
GN	MUC3.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OC	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=97476275; PubMed=9334251;	
RA	Gum J.R. Jr., Ho J.T.L., Pratt W.S., Hicks J.W., Hill A.S.,	
RA	Vinall L.E., Robertson A.M., Swallow D.M., Kim Y.S.;	
RT	"MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl	
RT	terminus, and a novel upstream repetitive region."	
RL	J. Biol. Chem. 272:26678-26686(1997).	
RL	termilus, and a novel upstream repetitive region."	
DR	ENBL; AF007194; AAC0272.1; -	
DR	ENBL; AF007196; AAB84383.1; -	
DR	ENBL; AF007195; AAB84382.1; -	
DR	InterPro: IPR000561; EGF-like.	
DR	SMART: SM00181; EGF; 1.	
DR	PROSITE: PS00022; EGF_1; UNKNOWN_2.	
DR	PROSITE; PS01186; EGF_2; 1.	
KW	EGF-like domain; Glycoprotein.	
FT	NON_TER	1
FT	SEQUENCE	957 AA; 99554 MW; E6772256D600A1F CRC64;

Qy	650	GGATGTGGCGCTCACAGGCGCACTTCAGCGTATTCCGGCTGTATCCGGTGTTACGCCGTCCGCACA	709
Db	285	GlyLeuProSerTrpValThrThrLysThrThrSerHisIleThrProGlyLeuThr	304
Qy	710	ACGGCCGGCGAATAACAGCAGAATAAACCCCTATTAAGCCGTTTTACTTCGCTGCCG	769
Db	305	SerSerIleThrThrThrGluThrThrSerHisSerThrProGlyPheThrSer-----	322
Qy	770	TATCATCGCAAGCTCAAAGCGTAGCTATCCATGACCATTGGTTTCCGGGCGAGTA	829
Db	323	--SerIleThrThrThrGluThrThrSerGluSerThrProSerLeuSerSerSerThr	341
Qy	830	CGGATCGCTACCTACTCTTGGGGTGTGGGAAGATCGTATCGAAACAGCAACATCCTC	889
Db	342	-----IleTySerThrValSerThrSerThrAlaIle	353
Qy	890	GTATCGAAGTTCGGGTATCAAGCAAGGCATCTGGAAGAACCATGAGCCGACGATCAGG	949
Db	354	ThrSerHisPheThrThrSerGluThrAlaValThrProThrProValThrProSerSer	373
Qy	950	CTACCCGTATCAATATGCCAGCAGTACTCTCAGAGTCTTAACCTATTGGAAGAATTCGA	1009
Db	374	LeuSer---ThrAspIleProThrSerLeuArgThrLeuThrProSer-----	389
Qy	1010	TCGTATGAACCCGGTCTCGCTTGCAGTGATAGTCTGAAGCGTG-----	1060
Db	390	SerValGlyThrSerThrSerLeuThrThrThrThrAspPheProSerIleProThrAsp	409
Qy	1061	-----CCGAGGAAGAGCAT-----TCGCAGACTGGATCGGTAAAGAACGCCA	1102
Db	410	IleSerThrLeuProThrArgThrHisIleIleSerSerSerProSerIleGlnSerThr	429
Qy	1103	AGAGTGTGTCATGGCGATGTATTGCTCTCTCGAAAAGCTTATAAGAAAGGACCCA	1162
Db	430	GluThrSerSerLeuValGlyThrThrThrSerProThrMetSerThrValArgMetThrLeu	449
Qy	1163	AGGCCAACCGTGAAGTACTTATTTGAGCGAGACGCTCTTCGGTGGTACCGAGGTGCTC	1222
Db	450	ArgIleThrGluAsnThrProfileSerSerPheSerThrSerIleVal-----	465
Qy	1223	GTTTTGCACAGTTTGGCAGCGCATTTGGCTACAAATCTCTGATGCTCATCCGCGGTATCTCA	1282
Db	466	ValIleProGluThrProThr-----GlnThrProProValLeuThr	479
Qy	1283	AATCGCTTGACGACAAGTACA-----AAGACTACC	1312
Db	480	SerAlaThrGlyThrGlnThrSerProAlaProThrThrValThrPheGlySerThrAsp	499
Qy	1313	TCGCCTCGCTCGACCCTAAGGTGCTCCCGCCATGCTCGATATTGTACGCCGGGTATCC	1372
Db	500	SerSerThrSerThrLeuHisThrLeuThrProSerThrAlaLeuSerThrIleValSer	519
Qy	1373	-----CTGCCG-----	1378
Db	520	ThrSerGlnValProIleProSerThrHisSerSerThrLeuGlnThrThrProSerThr	539
Qy	1379	-----ACAAGCTCCCAGATATATCAAGAATCTANTCCACAGA	1417
Db	540	ProSerLeuGlnThrSerLeuThrSerThrSerGluPheThrThrGluSerPheThrArg	559
Qy	1418	AATTCAAAGCGCACGAAGATGACAGCTTCGTATTCGACAAGAGTGTGTTCTC---	1474
Db	560	GlySerThrSerThr---AsnAlaIleLeuThrSerPheSerThrIleIleTrpSerSer	578
Qy	1475	-----CTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCATGGCAGGAAGAAAGTTTG	1528
Db	579	ThrProThrIleIleMetSerSerProSerSerAlaSerIleThr--ProValPheS	598
Qy	1529	CCAAGGCTATCGAAGNAGATCCGGCA-----	1554
Db	598	erThrThrIleHisSerValProSerSerProTyriIlePheSerThrGluAsnValGlys	618
Qy	1555	-----GTAGAGCTTTCCAAGAGCGGTAAATAGCTGCTCGCTCCGCGCTA	1594

```

Db 618 erAlaSerIleThrGlyPheProSerLeuSerSerAlaThrThrSerThrSer 638
Qy 1595 TTCAGCGCGATGCGATGCCAATGCTATGCAATGAGAGGGCAAGCGTCTTTTCTTTG 1654
Db 638 hrSerSerSerLeuThrThrAlaLeuThrGluIleThrPro-PheSerTyIleSerLeu 657
Qy 1655 CCGTTTTCGGTGGAGATGACCCCGGAGGTG-----CTCTCCGAGCGCATCCCACT 1705
Db 658 ProSerThrThrProCysProGlyThrIleThrIleThrIleValProAlaSerProthr 677
Qy 1706 TCACCATCGTA-----TGAGCTAGCGGTCCA----- 1732
Db 678 AspProCysValGluMetAspProSerThrGluAlaThrSerProProThrThrProLeu 697
Qy 1732 ----- 1732
Db 698 ThrValPheProPheThrThrGluMetValThrCysProThrSerIleSerIleGlnThr 717
Qy 1732 ----- 1732
Db 718 ThrLeuThrThrTyMetAspThrSerMetMetProGluSerGluSerSerIleSer 737
Qy 1733 -----TCAAGGATATGAACCGCAGGAGCGTGCCTGTGATCAACTATCATAGGACAG 1783
Db 738 ProAsnAlaSerSerThrGlyThrGlyThrValProThrAsnThrValPheThrSer 757
Qy 1784 GCAAGGGCGTATTGGAGAGCAGGATCTTAAGAGCGATGAGTTGCGGTACAGGAGAATA 1843
Db 758 ThrArg-----LeuPro-----Thr 762
Qy 1844 TCCTCGACCTCTCCGACCAAAACTATGTGCGTATGCGGAGAACG---GTCAGCTCC 1900
Db 763 SerGluThrThrLeuSerAsnSerSerValIleProLeuProLeuProGlyValSerThr 782
Qy 1901 ATATCGCTTCTTCATCGAACACGATCAGCGGGGTAACTCCGGTAGCCCGGTATTTCG 1960
Db 783 IleProLeuThrMetLysProSerSer-----LeuProThrIle 796
Qy 1961 ATAGAAGCGCGTGTGATCGTCTGCTTTCGATGGCACTGGGAGCTATGAGTGTG 2020
Db 797 LeuArgThrSerSerLysSerThrHisProSerProProThrThrArg----- 812
Qy 2021 ACATCGAGTTCGAACCCGATCTGCGACCAATCAGCGTGGACATCCGCTACGTTCTCT 2080
Db 813 ---ThrSerGluThrProValAlaThrThrGln-----ThrProThrThrLeuThr 828
Qy 2081 TCATGATTGACAATGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2125
Db 829 SerArgArgThrThrArgIleThrSerGlnMetThrThrThrGlnSer 843
RESULT 5
Q9UKW9 PRELIMINARY; PRT; 1217 AA.
AC Q9UKW9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Intestinal mucin 3 (Fragment).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE=99443732; PubMed=10512748;
RA Crawley S.C., Gum J.R. Jr., Hicks J.W., Pratt W.S., Aubert J.P.,
RA Swallow D.M., Kim Y.S.
RT "Genomic organization and structure of the 3' region of human MUC3;
RT alternative splicing predicts membrane-bound and soluble forms of the
RT mucin."

```

```

RL Biochem. Biophys. Res. Commun. 263:728-736(1999).
DR EMBL; AF113616; AAF13032.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00024; SEA; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
SQ SEQUENCE 1217 AA; 129171 MW; 078344003C969664 CRO64;

Alignment Scores:
Pred. No.: 0.0302 Length: 1217
Score: 157.00 Matches: 170
Percent Similarity: 33.00% Conservative: 93
Best Local Similarity: 21.33% Mismatches: 289
Query Match: 4.11% Indels: 245
Db: 4 Gaps: 32

US-10-008-355-1 (1-2139) x Q9UKW9 (1-1217)
Qy 155 ATTCGCTCTACAGTTTCGACAAAGCGTCCATTGCAATGCGGTATCTTCGGTGGCG 214
Db 154 IleThrThrThrGluThrThrSerHisAspThrPro-----SerPheThrSer 169
Qy 215 GATGTACCGGTATCACAGTGTCCGATCAGG-----GCCTGATCTTTACCAACC 262
Db 170 SerIleThrThrSerGluThrProSerHisSerThrProSerSerThrSerLeuThr 189
Qy 263 ACCATGCGGATAGCTGCTATCCAGAGCCAAAGCACGGTGGATCAGCACTATCTGCGCG 322
Db 190 ThrThrLysThrThrSerHisSerThrProSer-----PheThrSerSerIleThr 206
Qy 323 ATGTTTCTGTTCTCGACGATGGTGAGAGCTTCCGATTCGGGTCTTTCCGTGAAGT 382
Db 207 ThrThrGluThrThrSerHisSerAlaArgSerPhe----- 218
Qy 383 ATTCGCGCAAGATCGTGAAGGTAAAGGACAGGTAGAGGACAGCTCAAGGTATCACTG 442
Db 219 -----ThrSerSerIleThrThr 224
Qy 443 ACGAGATGAGCGTCTGCGCAAGGCTCAGGAGGTATGCCAAGAACTGCCCAAAAGAAA 502
Db 225 ThrGluThrThr-----SerHisAsnThrArgSerPheThrSerSerIle 239
Qy 503 ATGACAGCAGAGAACCAACTCT-----GCATCGTAGAGCCTTTCTATTCCAAACAG 553
Db 240 ThrThrThrGluThrAsnSerHisSerThrThrSerPheThrSerSerIleThrThr 259
Qy 554 AATACTTCTCTCATGCTACCATGTATTCAGGAGCTTCGTATGATGATTTGCTCTCCCA 613
Db 260 GluThrThrSerHisSerThr-----Pro 267
Qy 614 GCTCTGTAGTAGTAAAGTTCGAGCGGATACGACAACT----- 649
Db 268 SerPhe-----SerSerSerIleThrThrGluThrProLeuHisSerThrPro 284
Qy 650 GGATGTGCGCGGTACACAGCGCGGACTTTCAGCGTATTCGGGTGTATGCGGTGCGGACA 709
Db 285 GlyLeuThrSerTrpValThrThrLysThrThrSerHisIleThrProGlyLeuThr 304
Qy 710 ACCGCGCGCGGATACAGCAGGAGCAATAACCCCTATAAGCCCGTTTACTTCGTCGCGG 769
Db 305 SerSerIleThrThrGluThrThrSerHisSerThrProGlyPheThrSer----- 322
Qy 770 TATCCATCAAGGCTACAGGCTGACACTATGCCATGACCATCGTTCGCCGCGCAGTA 829
Db 323 ---SerIleThrThrThrGluThrThrSerGluSerThrProSerLeuSerSerThr 341
Qy 830 CGGATCGCTACTCTTCTTGGGGTGTGGAAGATCGTATCGAAAAACGAGAACAAATCCTC 889

```

```
Db 342 -----ileTySerThrValSerThrSerThrAlaIle 353
QY 890 GTATCGAAGTTCGGGTATCAACGAGGATCTGGAGGAGCCATGAGCGGATCAGG 949
Db 354 ThrSerHisPheThrThrSerGluThrAlaValThrProThrProValThrProSerSer 373
QY 950 CTACCGGTATCAATATGCGCAAGTATGCTCAGAGTCTAACTATTGGAAGAATTGCA 1009
Db 374 LeuSer-----ThrasPleProThrThrSerLeuArgThrLeuThrProSer----- 389
QY 1010 TCGGTATGAACCGCGTCTCGCTGCTTGGAGTGATAGTGTGTAAGCGTG----- 1060
Db 390 SerValcylThrSerThrSerLeuThrThrThrThrAspPheProSerIleProThrAsp 409
QY 1061 -----CCGAGGAAGAGCAT-----TCGCACACTGGATCCCTAAGACGGCA 1102
Db 410 IleSerThrLeuProThrArgThrHisIleIleSerSerProSerIleGlnSerThr 429
QY 1103 AGAGTCTGCTATGGCGATGATTCTCTCTCGAAAAGGCTTATAAGGAGGAGGCA 1162
Db 430 GluThrSerSerLeuValGlyThrThrSerProThrMetSerThrValArgMetThrLeu 449
QY 1163 AGCCCAACCGTGAGTATTATTGAGCGAGACGCTCTTCGGTGATACCGAGGTGGTTC 1222
Db 450 ArgIleThrGluAsnThrProIleSerSerPheSerThrSerIleVal----- 465
QY 1223 GTTTTGCACAGTTTCCCAAGCATGGCTACAAATCCTGATGCTATGCCGGTATCCCTCA 1282
Db 466 ValIleProGluThrProThr-----GlnThrProProValLeuThr 479
QY 1283 AATCGCTTGACGACAGTAGTACA-----AAGACTACC 1312
Db 480 SerAlaThrGlyThrGlnThrSerProAlaProThrThrValThrPheCylSerThrAsp 499
QY 1313 TCCCTCGCTCGACCGTAAGTGCTGCCCGCATGCTCGATATGTCACCGCGCGTATCC 1372
Db 500 SerSerThrSerThrLeuHisThrLeuThrProSerThrAlaLeuSerThrIleValSer 519
QY 1373 -----CTGCCG----- 1378
Db 520 ThrSerGlnValProIleProSerThrHisSerSerThrLeuGlnThrThrProSerThr 539
QY 1379 -----ACAAGTCCCGGATATATTAAAGATGTAATCGACAAGA 1417
Db 540 ProSerLeuGlnThrSerLeuThrSerThrSerGluPheThrThrGluSerPheThrArg 559
QY 1418 AATTCAAGCGGACAGGAAGTATGCGAGCTTCGTATTCGACAAGATGTGGTTC----- 1474
Db 560 GlySerThrSerThr-----AsnAlaIleLeuThrSerPheSerThrIleIleTrpSerSer 578
QY 1475 -----CTTATAGGACAAAGTTCCATGCCATGCTCAAGTCCATGGACAGGAAAGTTTG 1528
Db 579 ThrProThrIleIleMetSerSerSerProSerSerAlaSerIleThr--ProValPheS 598
QY 1529 CCAAGCTATCGAAGAAAGTCCGGCA----- 1554
Db 598 erThrThrIleHisSerValProSerSerProTyriIlePheSerThrGluAsnValGlyS 618
QY 1555 -----GTAGAGCTTTTCCAAGCGCTAATAGCTGCTGCTCGCGTA 1594
Db 618 erAlaSerIleThrGlyPheProSerLeuSerSerSerAlaThrThrSerThrSerSerT 638
QY 1595 TTCAGCGCGATGCGATGGCCATGCTATGCCATTGACAGAGGCGCAGGCTCTTTTCTTG 1654
Db 638 hrSerSerSerLeuThrThrAlaLeuThrGluIleThrPro-PheSerTyriIleSerLeu 657
QY 1655 CGGGTTTGGTGATGTACCCCGGACGTG-----CTGCGCGAGCGCATGCCAAT 1705
Db 658 ProSerThrThrProCysProGlyThrIleThrIleThrIleValProAlaSerProThr 677
QY 1706 TCACCATGCGTA-----TGAGTACGGTCCA----- 1732
|||||
```

```
Db 678 AspProCysValGluMetAspProSerThrGluAlaThrSerProProThrThrProLeu 697
QY 1732 ----- 1732
Db 698 ThrValPheProPheThrThrGluMetValThrCysProThrSerIleSerIleGlnThr 717
QY 1732 ----- 1732
Db 718 ThrLeuThrThrTyThrMetAspThrSerSerMetMetProGluSerGluSerSerIleSer 737
QY 1733 -----TCAAGGATATGAACCGCAGGAGCGTCCCTGGTACAACTATCATACGACAG 1783
Db 738 ProAsnAlaSerSerSerThrGlyThrGlyThrValProThrAsnThrValPheThrSer 757
QY 1784 GCAAGCGCGTATTGGAGAACGAGGATCCTAAGACGATGAGTTTCCCGTACAGGAGAATA 1843
Db 758 ThrArg-----LeuPro-----Thr 762
QY 1844 TCCTCGACCTCTTCGCGACCAAAACTATGGTCGTATATCCCGAGAAGG---GTCAGCTCC 1900
Db 763 SerGluThrThrTrpLeuSerAsnSerSerValIleProLeuProLeuProGlyValSerThr 782
QY 1901 ATATCGCTTCTCTATCGAACACGACATACCGGCGGTAACTCCGGTACGCCCGTATTTCG 1960
Db 783 IleProLeuThrMetLysProSerSer-----LeuProThrIle 796
QY 1961 ATAAGAAGCGCGTCTGATCGTCTTGTTCGATGGCAACTCGGACATCGCTACGTTCTCT 2080
Db 797 LeuArgThrSerSerLysSerThrHisProSerProProThrThrArg----- 812
QY 2021 ACATCGAGTTTCAACCGCATCTCGACGCACATCAGCGTGGACATCGCTACGTTCTCT 2080
Db 813 ---ThrSerGluThrProValAlaThrThrGln-----ThrProThrThrLeuThr 828.
QY 2081 TCATGATTGACAAATGGGTCAGTCCCGCTCTCATCCAAGAGC 2125
Db 829 SerArgArgThrThrArgIleThrSerGlnMetThrThrGlnSer 843
RESULT 6
Q8TDH7 PRELIMINARY: PRT; 1029 AA.
ID Q8TDH7 AC Q8TDH7;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Intestinal membrane mucin MUC17 (Fragment).
GN MUC17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21845432; PubMed=11855812;
RT Gum J.R., Jr., Crawley S.C., Hicks J.W., Szymkowski D.E., Kim Y.S.;
RA "MUC17, a novel membrane-tethered mucin.";
RL Biochem. Biophys. Res. Commun. 291:466-475(2002).
DR EMBL; AF430017; AAL89737.1; -.
FT NON_TER
SQ SEQUENCE 1029 AA; 108320 MW; 5E1AD19E4BF948E3 CRC64;
Alignment Scores:
Pred. No.: 0.0301 Length: 1029
Score: 156.50 Matches: 147
Percent Similarity: 32.56% Conservative: 63
Best Local Similarity: 22.79% Mismatches: 230
Query Match: 4.10% Indels: 205
Gaps: 29
US-10-008-355-1 (1-2139) x Q8TDH7 (1-1029)
QY 536 CTTTCTATTCCAACGAAGTACTTCTCTCATCGTCTACGATGATTATCAAGGACGTTTCGTA 595
||||| J||||| :||||| :|||||
```

```
Db 104 MetSerThrProSerGluGlySerSerSerLeuThrThrMet----- 117
QY 596 TGGTATTGCTCCTCCAGCTCTGTAGTAAAGTTTCGGAGCGATACGACAACTGGATGT 655
Db 118 -----LeuLeuSerSerThrTyrValThrSerSerGluAlaSerThrProSerThr 134
QY 656 GCGCGCTCACCGGGGACT----- 676
Db 135 ProSerValAspArgSerThrProValThrThrSerThrGlnSerAsnSerThrProThr 154
QY 677 -----TTCAGCGTATTCCGCG 691
Db 155 ProProGluValIleThrLeuProMetSerThrProSerGluValSerThrProLeuThr 174
QY 692 TGTATCGCGTCCGCGCAAC-----GGCGCGCGAATACAGCAAGGACAATA 739
Db 175 IleMetProValSerThrThrSerValThrIleSerGluAlaGlyThrAlaSerThrLeu 194
QY 740 -----AACCTATAAGC 751
Db 195 ProValAspThrSerThrProValIleThrSerThrGlnValSerSerProValThr 214
QY 752 CCG---TTTACTTCGCTCCGCTAT-----CCATGCAAG---GCTACAAGGCTCAGCACT 799
Db 215 ProGluGlyThrThrMetProIleThrProSerGluGlySerThrProLeuThrThr 234
QY 800 ATGCCATGACCATCGGTTTCCCGGGCAGTACGGATCCTACCTCCTCTTGGGGTGG 859
Db 235 MetPro---ValSerThrThrArgValThrSerSerGluGlySerThrLeuSerThrPro 253
QY 860 AAGATCGTATCGAAGAACAGAACATCTCGTATCGAAGTTCCGGGTATCAAGCAAGCA 919
Db 254 SerValValThrSerThr---ProValThrThrSerThrGluAlaIleSerSerAla 272
QY 920 TCTGGAAGGAGCCATGAGCGCAGATCAGCTACCGCTACCGTCAATATATGCGCAGCAAGTATG 979
Db 273 Thr-----LeuAspSerThrThrMetSerValSerMet 283
QY 980 ---CTCAGAGTGTACTATTGGAAGAAATCGATCGGTATGAACCGGGTCTCGCTCGTC 1036
Db 284 ProMetGluIleSerThrLeuGly-----ThrThrIleLeuVal 296
QY 1037 TTGAGTGTAGTGTCTGAAGCGTCCGAGGAGGAGGATTCGGAGACTGGATCCGTAGA 1096
Db 297 SerThrThrProValThrArgPheProGluSerSerThrProSerIleProSerValTyr 316
QY 1097 AGCGCAAGAGTGTCTGTATGGCGCATCTATTGTCTCTCTCGAAAGGCTTATAAGGAAG 1156
Db 317 ThrSer-----MetSerMetThrThrAlaSerGluGlySerSerProThrThrLeu 334
QY 1157 GAGCAAGGCCAACCGTACAGTACTATTGACGAGAGCGCTCTCGGTGGTACCGAGG 1216
Db 335 GluGlyThrThrMetProMetSerThrThrSerGluArgSerThrLeu----- 351
QY 1217 TGGTTCGTTTGCACAGTTGCCAACCGATTCGGTACAAATCTGTATGCTCATCCGGTA 1276
Db 352 -----LeuThrThrValLeuIleSerProIle 360
QY 1277 TCC-----TCAAAAT 1285
Db 361 SerValMetSerProSerGluAlaSerThrLeuSerThrProProGlyAspThrSerThr 380
QY 1286 CGCTTGACGACAGTACAAGACT----- 1309
Db 381 ProLeuLeuThrSerThrLysAlaGlySerPheSerIleProAlaGluValThrThrIle 400
QY 1310 -----ACCTCCCTCGCTCGACCGTAAGGTGCTCCCGCCATGCTCGCATATTG 1357
Db 401 ArgIleSerIleThrSerGluArgSerThr-----ProLeuThrThrLeu 415
QY 1358 TACCGCGCGGATCCCTGCGCAGCAAGTCCCGCATATATTCAAGAATGTAATCCGACAAGA 1417
Db 416 LeuValSerThrThrLeuProThrSerPhePro-----Gly 427
```

```
QY 1418 AATTCAAAGCGCAGACAGAAAGTATATGACAGCTTCGTATTTCGACAGAGTGTGGTTCCTT 1477
Db 428 AlaSerIleAlaSerThrProProLeuAspThrSer----- 439
QY 1478 ATACCGACAACTTCCATGCCATGCTCAAGTCCATCGGACAAAGGAAAGTTTCCCAAGCTA 1537
Db 440 -----ThrThrPheThrProSerThrAspThrAlaSerThrProThrIleProValAla 457
QY 1538 TCG-----AGAAAGATCCGCGAGTAGAGCTTTTCAAGAGCGCTAATAGCTGCTG 1585
Db 458 ThrThrIleSerValSerValIleThrGluGlySerThrProGlyThr----- 473
QY 1586 CTCGCGCTATTACGCGCGATGCGATGCCAATGCTATGCGCATTCGAGAGGGCAAGGCTC 1645
Db 474 ---ThrIlePheIlePro-----SerThrProValThrSerSerThrAlaAspVal 489
QY 1646 TTT-----TCTTTGCCGTTTCCGTGAGATCT-----ACC 1675
Db 490 PheProAlaThrThrGlyAlaValSerThrProValIleThrSerThrGluLeuAsnThr 509
QY 1676 CCGGACGTGCTCTGCCGAGCGATGCCAACTCA----- 1708
Db 510 ProSerThrSerSerSerSerThrThrThrPheSerThrThrLysGluPheThrThr 529
QY 1709 CCATGCGTATGAGTACGGTCCATCAAGGATATGAACCGCAGGAGCGTCCCTGGTACA 1768
Db 530 ProAlaMetThrThrAlaAlaProLeuThrTyrValThrMetSerThrAlaProSerThr 549
QY 1769 ACTATCATACGACAGCGAAGGGCGTATTGGAGACGAGGATCTTAAGAGCGATGAGTTTG 1828
Db 550 ProArgThrThrSerArgGlyCysThrThrSerAlaSerThrLeuSerAlaThrSerThr 569
QY 1829 CCGTACAGGAGAAATA-----TCTCGACCTCTTCCGCA 1861
Db 570 ProIleThrSerThrSerValThrThrArgProValThrProSerSerGluSerSerArg 589
QY 1862 CCAAAAACTATGTCGCT-----ATCCGAGAACGGTCACTCCATATCCGTTTCTCTAT 1915
Db 590 ProSerThrIleThrSerHisThrIleProProThrPheProAlaHisSerSerThr 609
QY 1916 CGAACACGACATCACGGCGGTAACTCCGGTAGCCCGGATTCGATTAAGAACGCGCGTC 1975
Db 610 ProProThrThrSerAlaSerSerThrThrValAsnProGluAlaValThrThrMetThr 629
QY 1976 TGAATCGTCTGCTTTCGATGGCAACTGGGAAGCTATGAGTGGTGCATCGAGTTTCAAC 2035
Db 630 ThrArgThrLysProSerThrArgThrThrSerPheProThrValThrThrAlaVal 649
QY 2036 CCGATCTGACGCGCACAAATCAGCGTGGACATPCGCTACGTTCTCTTCATGATTCACAAT 2095
Db 650 Pro-----ThrAsnThrThrIleLysSer-----Asn 658
QY 2096 GGGTCACTGCCCCC 2110
Db 659 ProThrSerThrPro 663
```

## RESULT 7

```
QY 09UQ36 PRELIMINARY; PRT; 1275 AA.
ID 09UQ36
AC 09UQ36;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RNA binding protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ontaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, partial cds.";
```



```

QY 1830 CGTACAGGAGAAATATCTCGACCTCTTCCGACCAACAAATATGCTCGCTATGCCGAGAA 1889
DB 617 roAlaThrArgAsnHisSerGlySerArgThrPro-----ProVala 631
QY 1890 CGGTCACTCCATATCGCTTTCCTATCGAACACGACATCACGGCGGTAACTCCGGTAG 1949
DB 631 laLeuAnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA 651
QY 1950 CCCCGTATTCGATAAGAACGCCCTCTCTGATCGGTCTTTCGATGGCAACTGGGAAGC 2009
DB 651 sp-----ArgCysArgSerProGlyMetL 659
QY 2010 TATGAGTGGTGACATCGAGTTCGAACCCGATCTCGACGCGACATCAAGCGTGGG----- 2063
DB 659 euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer 678
QY 2064 -----CATCCGCTACGTTCTCTCATCATGATGACAAATGGGGTCAAGCCCGCTCT 2114
DB 679 MetMetAspGlyProGlyProArgGlyProAspHisGlnArgThrSerValProGluAsn 698
QY 2115 CATCCAAGAGCTGAAGTTGATCTA 2138
DB 699 HisAlaGlnSerArgIleAlaLeu 706

RESULT 8
O15038 PRELIMINARY; PRT; 1783 AA.
AC O15038;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0324 protein (Fragment).
GN KIAA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=BRAIN;
RX MEDLINE=9734984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002322; BAA20782.2; -.
FT NON_TER 1
SQ SEQUENCE 1783 AA; 190940 MW; 660302f6FD4179AB CRC64;

Alignment Scores:
Pred. No.: 0.043 Length: 1783
Score: 156.00 Matches: 184
Percent Similarity: 34.42% Conservative: 81
Best Local Similarity: 23.90% Mismatches: 276
Query Match: 4.08% Indels: 229
DB: 4 Gaps: 35

US-10-008-355-1 (1-2139) x O15038 (1-1783)
QY 27 TCTCGGACAGCCCTGCTGTGTTGGTGTTCAGGGGTAGCCCAAGCCGACAAAGCATGTG 86
DB 610 SerGlySerSerProGluVal-----AspSerLysSerArg-----Leu 622
QY 87 GCTCCTCAACCAACTCAATCAGAGAGATCTGGATCGATCGGTGAGCTCGGCTTTAGCT 146
DB 623 SerProArgArgSerArgSerGlySerSerProGluValLysAspLysProArgAlaAla 642
QY 147 CCCGTTGGATTCGCTACAGTTTCGACAGCCGCTCCCATGCCCGGTGTTATCTT 206
DB 643 Pro----- 643

```

```

QY 207 CGTGGCGGATGTACCGGTATCAGAGTGTCCGATCAGGGGCTGATCTTTTACCACACCA 266
DB 644 -----ArgAlaGlnSerGlySerAsp----- 650
QY 267 CTGGGATACGGTGTATTCAGAGCCCAAGCAGCGGTGGATCAGACTATCTGCGCGGATGG 326
DB 651 -----SerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 666
QY 327 TTTCTGCT-----TTCCTCGCAGCATGGGTGAGGAGCTTCCGATTCGGGTCTTTCCGTGAA 380
DB 667 SerArgSerGlySerSerLysGlyArgGlyProSerProGluGlySerSerSerThr 686
QY 381 GTATCTCCGAAGATCTGGAAGGTACGGACAGGTAGAGGACAGCTCAAGGGTATCAC 440
DB 687 GluSerSerProGluHisProProLysSer-----ArgThrAlaArg----- 700
QY 441 TGACGAGATGGAGGCTGTCCGCAAGCTCAGGAGGTATGCCAAGACT---GGCCAAAAA 497
DB 701 ---ArgGlySerArgSerSerProGluProLysThrLysSerArgThrProProArgArg 719
QY 498 AGAAATGCAGACGAGAACCAACTCTCATCTAGAGCTTTCTTATTCACAAACA----- 551
DB 720 ArgSerSerArgSerProGluLeuThrArgLysAlaArgLeuSerArgArgSerArg 739
QY 552 -----CGAATACTTCTCT---CATCGTCTACGATGT 578
DB 740 SerAlaSerSerProGluThrArgSerArgThrProProArgHisArg----- 756
QY 579 ATTCAGGAGCGTTCGTATGGTATTTGCTCC-----TCCAGACTCTGTAGTAAAGTT 629
DB 757 -----ArgSerProSerValSerSerProGluProAlaGluLysSerArgSerSer 773
QY 630 CGGAGGCGATACGACAACTGGATGTGGCGCGTCACAGGGCGACATTACCGGTATTCG 689
DB 774 ArgArgArgArg-----SerAlaSerSerProArgThrLysThrSer 788
QY 690 CGTGTATCCCGGTCCGACAAACCGCGCGGTATACAGCAAGGACAAATAAACCTATAA 749
DB 789 ArgArgGlyArgSerProSerProLysProArg----- 799
QY 750 GCCCGTTTACTTCCTCGCTATCCATCCAGGCTACAGGCTCAGGACTATGCCATGAC 809
DB 800 ---GlyLeuGlnArgSerArgSerArgSerArgGluLys----- 812
QY 810 CATCGGTTTCCCGGCGAGTACGGATCGTACCTCCTCTTCTGGGGTGTGGAAGATCGTAT 869
DB 813 -----ThrArgThrThr 816
QY 870 CGAAACGAGAACAACTCCTCGTATCGAAGTTTCGGGTATCAAGCAAGGCACTCTGGAAGA 929
DB 817 ArgArgArgArgSerGly-----SerSerGlnSerThrSerArgArgGlnArg 834
QY 930 AGCCATGAGCGCAGATCAGGCTACCGTATCAAA-----TATGCCAGCAA 974
DB 835 SerArgSerArg-SerArgValThrArgArgArgArgGlySerGlyThrHisSerArg 854
QY 975 GTATGCTCAGAGTCTAACTATTGGAAGAAATTCGATCGGTATGAACCGCGCTCTCGTCG 1034
DB 854 gSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArgGlyArgSerArg 874
QY 1035 TCTTTGACGTGATAGTGTGAAGCTGCCGAGGAAAGAGCATTC---GCAGACTGGATCCG 1091
DB 874 gThrProProThrSerArgLysArgSerArgSerArgThrSerProAlaProLysArg 894
QY 1092 TAAACAGCGCAAGAGTCT-----CTGCTCTATGGCGATGTATTGCTTCT 1133
DB 894 gSerArg-SerArgAlaSerProAlaThrHisArgSerArgSerArgSerArgThrProLeuI 914
QY 1134 TCTCGAAAGGCTTATAGGAAGGAGGAGCCCAAGCCGTCAGATGACTTATTTTTCAGGGA 1193
DB 914 leSerArgArgArgSerArgSerArgThrSerProValSerArgArgSerArgSerArg 934

```



```
Qy 1194 CA-----CCGCTTTCGGTGGTACCGAGGTGGTTTCG 1223
||
Dy 934 rgThrSerValThrArgArgSerArgSerArgAlaSerProValThrArgArgS 954
1224 TTTGTCACAGTTTGCACGATGGCTGACAAATCCTGATGCTCATGCCGATCCTCAA 1283
||
Dy 954 erArgSerArgThrPro-----ProValThrArgA 964
1284 ATCGCTTGACGACAGTAAAGACTACCTCCCTCGCTCGACCGTAAGGTGCTGCCGC 1343
||
Dy 964 rgArgSerArgThrProThrThrThrArgArgSerArgSerArg-----ThrPro 983
1344 CATGCTCGATATTGACCGCGCGCTATCCCTCGCG-----1378
983 roValThrArgArgSerArgSerArgThrProProValThrArgArgSerArgS 1003
1379 --ACAGCTCCCGATATATTCAGATGTAATCGCAAGAAATTCAGGCGCACAGAA 1436
||
Dy 1003 erArgThrSerProFileThrArgArgSerArgSerArgThrSerProValThrArgA 1023
1437 GA-----AGTATCGAGACTTCGTATTCGACAAAGAGTGTGCTTCTATACGCAAGTT 1490
1023 rgArgSerArgSerArgThrSerProValThrArg-----ArgArgS 1037
1491 CCATGCCATGCTCAAGTCCATGGACAAG-----AAAAGTTTGCCAAAGC 1535
||
Dy 1037 erArgSerArgThrSerProValThrArgArgSerArgSerArgThrProProAlaI 1057
1536 TATCGAGAAAGATCGGACAGTAGAGC-----TTTCCAAAGCGTATAGTGTGCTGCTG 1589
1057 leArgArgSerArgSerArgThrProLeuLeuProArg-----1070
1590 CGCTATTCCAGCGCATGGTGGCAATGCTATGCCATTCAGAAAGGCAAGCGTCTTTT 1649
1071 -----LysArgSerArgSerArgSerProLeuAlaIleArgArgSerArgSerA 1088
1650 CTTTCCCGTTTGGGTGAGATGTACCCCGGAGCGTCTCTGCCGAGCGATGCAACTTCAC 1709
1088 rgThrProArgThrAlaArgGlyLys-----ArgSerLeuThrArgSerP 1103
1710 CATCGGTATGACTACGCTCCATCAGGATATGACCGGAGGAGCGTCTGCTGTACAA 1769
1103 ro-----ProAla-----IleArgArgArgSerAlaSerGlySers 1115
1770 CTATCATACGACAGCAAGGCGTATTGGAGAACAGCAGGATCTTAAGACGATGATTTGC 1829
1115 erSerAspArgSer-----ArgSerAlaThrProp 1125
1830 CGTACAGGAGATATCTTCGACCTTTCCCGACCAAAACTATGCTGCTATGCCGAGAA 1889
1125 roAlaThrArgAsnHisSerGlySerArgThrPro-----ProValA 1139
1890 CGGTACGCTCATATCGCTTTCCCTATCGAACACGACATACGCGGCGTAACTCGGTAG 1949
1139 laLeuAsnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA 1159
1950 CCCGCTATTTCGATAAAGACCGCTCTGATCGGTCTTCTTCGATGGCAACTGGGAAGC 2009
1159 sp-----ArgCysArgSerProGlyMetL 1167
2010 TATGAGTGGTGCATCGAGTTGCAACCCGATCTCGACGCGCAACTCAGCGTGA-----2063
1167 euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer 1186
2064 -----CATCCGCTACGTTCTCTCATGATTCGACAAATGGGGTATGTCGCCCGCTCT 2114
1187 MetMetAspGlyProGlyProArgGlyProAspHisGlnArgThrSerValProGluAsn 1206
2115 CATCCAAGCTGAGTGTGATCTA 2138
1207 HisAlaGlnSerArgIleAlaLeu 1214
RESULT 9
```

```
O60382
ID AC O60382 PRELIMINARY; PRT; 1791 AA.
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0324 (Fragment).
GN KIAA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT "Sequencing of Human Chromosome 16p13.3.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ricke D.O.;
RT "Large Scale Sequence Analysis and Annotation with the Sequence
RT Comparison Analysis (SCAN) System.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC004493; AAC08453.1;
FT NON_TER
SQ SEQUENCE 1791 AA; 191306 MW; 3A7B5530AE95F3E CRC64;

Alignment Scores:
Pred. No.: 0.0431 Length: 1791
Score: 156.00 Matches: 184
Percent Similarity: 34.42% Conservative: 81
Best Local Similarity: 23.90% Mismatches: 276
Query Match: 4.08% Indels: 229
DB: Gaps: 35

US-10-008-355-1 (1-2139) x O60382 (1-1791)
Qy 27 TCTCGGAGACGCTCGTGTGGTGTCTTACGGGTAGCCAAAGCCGACAAAGCATGTG 86
||| ||||| ||| ||||| ||| ||||| |||
Db 611 SerGlySerSerProGluVal-----AspSerLysSerArg-----Leu 623
Qy 87 GCTCCTCAACCAACTCAATCAGAGAAATCTGGATCGAATGCTGAGCTCGGCTTTAGCT 146
||| ||||| ||||| ||||| ||| ||||| |||
Db 624 SerProArgArgSerArgSerGlySerSerProGluValLysAspLysProArgAlaAla 643
Qy 147 CCCGTTGGATTTCGCTCTACAGTTTTCGACAGCCGCTCCATTGCCAATGCCGTGTATCTT 206
||| ||| Pro-----644
Qy 207 CGGTGGCGGATGTACCGGTATCAGTGTCTCGATGTCAGGCGCTGATCTTTACCAACCA 266
||| ||| ||| ||| ||| ||| ||| |||
Db 645 -----ArgAlaGlnSerGlySerAsp-----651
Qy 267 CTGCGGATACGGTCTATCCAGAGCAAAAGCACGGTGTGATCACCACCTATCTCGCGCATGG 326
||| ||||| ||||| ||||| ||| ||||| |||
Db 652 -----SerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 667
Qy 327 TTTTCGT-----TTCTCGCAGCATGGGTGAGAGCTTCCGATTCGGGTCTTTCCCGTCAA 380
||| ||| ||| ||| ||| ||| ||| |||
Db 668 SerArgSerGlySerSerLysGlyArgGlyProSerProGluGlySerSerSerThr 687
381 GTATCTCGGCAAGATCGTGAAGGTAAAGGACAGGTAGAAGGACAGCTCAGGGGTATCAC 440
||| ||| ||| ||| ||| ||| ||| |||
Db 688 GluSerSerProGluHisProProLysSer-----ArgThrAlaArg-----701
441 TGACGAGATGGAGCGTCTCGCAAAAGCTCAGGAGGTATGCCAAGAAGCT---GGCCAAACAA 497
||| ||| ||| ||| ||| ||| ||| |||
Db 702 ---ArgGlySerArgSerSerProGluProLysThrLysSerArgThrProProArgArg 720
498 AGAAATGCAGACAGAGAACCAACTCTGCATGCTAGAGCCTTTCTATTCCAAACAA-----551
```

Db	1024	rgArgSerArgSerArgThrSerProValThrArg	-----ArgArgS	1038
Qy	1491	CCATGCCATGCTCAAGTCCATGACCAAG	-----AAAAGTTGCCAAGGC	1535
Db	1038	erArgSerArgThrSerProValThrArgArgSerArgSerArgThrProAlaI	1058	
Qy	1536	TATCGAGAAATCCGGCAGTAGAC	-----TTTCCAAGACGGTAATAGCTGCTGCCTGC	1599
Db	1058	leArgArgArgSerArgThrProLeuLeuProArg	-----	1071
Qy	1590	CGCTATTTCAGSCCGATCGCATGCCAATGCCATGCTATGCATTGAGAGGGCAAGCGTCTTTT	1649	
Db	1072	-----LysArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSerA	1089	
Qy	1650	CTTTGCCGGTTTCGGTCAGATGATACCCCGAGCTGCTCTGCCAGCGATGCCAATCTCAC	1709	
Db	1089	rgThrProArgThrAlaArgGlyLys	-----ArgSerLeuThrArgSerP	1104
Qy	1710	CATCGGTATGAGTACGGTCCATCAAGGATATGAAACCGCAGAGCGGTGCTGTTGATACA	1769	
Db	1104	ro-----ProAla-----IleArgArgArgSerAlaSerGlySers	1116	
Qy	1770	CTATCATACGACAGCGCAAGGGCGTATTGGAGAGCAGGATCCTTAAGACGATGAGCTTTGC	1829	
Db	1116	erSerAspArgSer	-----ArgSerAlaThrProp	1126
Qy	1830	CGTACAGAGAATATCCTCGACCTCTTCGCACACCAAAACTATGCTCGCTATGCCGAGAA	1889	
Db	1126	roAlaThrArgAsnHisSerGlySerArgThrPro	-----ProVala	1140
Qy	1890	CGGTACGCTCCATCATGCTTCCTTCATCGAACACAGACATCACGGCGGTAACTCCGGTAG	1949	
Db	1140	laLeuAsnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA	1166	
Qy	1950	CCCCGTATTTCGATTAAGAACCGCGCTGATCGTCTTCGATGGCACTGGGAAGC	2009	
Db	1160	sp-----ArgCysArgSerProGlyMetL	1168	
Qy	2010	TATGAGTGGTGACATCGAGTTCGAACCCGATCTGCAGCGCACAACTCAGCGTGA	2063	
Db	1168	euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer	1187	
Qy	2064	-----CATCGCTACGTTCTCTTCATGATGTGAAATGGGTGTCAGTGCCTCCGCTCT	2114	
Db	1188	MetMetAspGlyProGlyProArgIleProAspHisGlnArgThrSerValProGluAsn	1207	
Qy	2115	CATCCAAGAGCTGAGTTGATCTA	2138	
Db	1208	HisAlaGlnSerArgIleAlaLeu	1215	
RESULT 10				
Q9UQ35				
ID	Q9UQ35	PRELIMINARY;	PRT; 2752 AA.	
AC	Q9UQ35;			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	RNA binding protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Okazaki S.; Umeki K.; Sawada Y.;			
RT	"Homo sapiens mRNA for RNA binding protein, complete cds.";			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AB016092; BAA83718.1; -			
DR	InterPro; IPR002965; P-rich_extensn.			
DR	PRINTS; PR01217; PRICHEXTENS.			
SQ	SEQUENCE 2752 AA; 299672 MW; 109C64F181097123 CRC64;			

Alignment Scores:

Pred. No.: 0.0535 Length: 2752  
 Score: 156.00 Matches: 184  
 Percent Similarity: 34.42% Conservative: 81  
 Best Local Similarity: 23.90% Mismatches: 276  
 Query Match: 4.08% Indels: 229  
 DB: 4 Gaps: 35

US-10-008-355-1 (1-2139) x 09U035 (1-2752)

Qy 27 TCTCGGAGCAGCTGCTGTTGGGTGCTTACAGGGGTACCCAAAGCCGACAAAGGCATGTG 86  
 Db 1579 SerGlySerProGluVal- - - - - AspSerLysSerArg- - - - - Leu 1591  
 Qy 87 GCTCCTCAACGAACATCAATCAGGAGAATCTGGATCGAATGCGTGAGCTCGGCTTTACGCT 146  
 Db 1592 SerProArgArgSerArgSerGlySerProGluValLysAspLysProArgAlaAla 1611  
 Qy 147 CCCGTTGGATTGCTCTACAGTTTCGACAGCCGTCATGCCAATGCCGTTATCTT 206  
 Db 1612 Pro- - - - - 1612  
 Qy 207 CGGTGGCGGATGACCGGTATCAGGTTCGATCCGATCAGGGCTGATCTTTACCAACCACCA 266  
 Db 1613 - - - - - ArgAlaGlnSerGlySerAsp- - - - - 1619  
 Qy 267 CTGCGGATACGGTGTATCCAGAGCCAAAGCAGCGGTGGATCAGCACTATCTCGCGATGG 326  
 Db 1620 - - - - - SerSerProGluProLysAlaProAlaProArgAlaLeuProArgArg 1635  
 Qy 327 TTTGCT- - - - - TTTCTGCGAGATGGGTAGAGCTTCCGATTCGGGTCTTTCCGTA 380  
 Db 1636 SerArgSerGlySerSerLysGlyArgGlyProSerProGluGlySerSerSerThr 1655  
 Qy 381 GTATCTCGCAAGATCGTGAAGGTACGGACAAGGTACAGGACAGCTCAAGGTTATCAC 440  
 Db 1656 GluSerSerProGluHisProProLysSer- - - - - ArgThrAlaArg- - - - - 1669  
 Qy 441 TGAGGAGATGGAGGCTGTCGCCAAAGCTCAGGAGGTATGCCAAAGAACT- - - GGCCAAAAA 497  
 Db 1670 - - - - - ArgGlySerArgSerSerProGluProLysThrLysSerArgThrProProArgArg 1688  
 Qy 498 AGAAATGCACAGACAGAACCAACTCTGCTGATCGTAGAGCTTTCTATCCACAA- - - - - 551  
 Db 1689 ArgSerArgSerSerProGluLeuThrArgLysAlaArgLeuSerArgArgSerArg 1708  
 Qy 552 - - - - - CGAATACTTCCT- - - - - CATCGTCTACGATGT 578  
 Db 1709 SerAlaSerSerSerProGluThrArgSerArgThrProProArgHisArg- - - - - 1725  
 Qy 579 ATTCAGGACGTTGCTATGTTGTTGCTC- - - - - TCCAGCTCTGTAGGTAAGTT 629  
 Db 1726 - - - - - ArgSerProSerValSerSerProGluProAlaGluLysSerArgSerSer 1742  
 Qy 630 CGGAGGCGATACGACAACTGGATGTGCGCCGCTACACGGGCGACTTCACGCTATTCG 689  
 Db 1743 ArgArgArgArg- - - - - SerAlaSerSerProArgThrLysThrThrSer 1757  
 Qy 690 CGTGTATGCGGTCCGACAAACCGCGCGGCGAATACAGCAAGGACAATAAACCTATAA 749  
 Db 1758 ArgArgGlyArgSerProSerProLysProArg- - - - - 1768  
 Qy 750 GCCGTTTACTTCGCTGCGGTATCCATGCAAGGCTACAAAGCTGACGACTATGCCATGAC 809  
 Db 1769 - - - - - GlyLeuGlnArgSerArgSerArgArgGluLys- - - - - 1781  
 Qy 810 CATCGTTTCCGGGAGTACGGATCGCTACCTCCTCTTTGGGGGTGGGAAGATCGTAT 869  
 Db 1782 - - - - - ThrArgThrThr 1785  
 Qy 870 CGAAACGAGAACAACTCTCGTATCGAAGTTCCGGGTATCAAGCAAGCATCTCGAAGGA 929  
 Db 1786 ArgArgArgAspArgSerGly- - - - - SerSerGlnSerThrSerArgArgGlnArg 1803

Qy 930 AGCCATGAGCGCAGATCAGGCTACCGCTATACAA- - - - - TATGCCACAA 974  
 Db 1804 SerArgSerArg- SerArgValThrArgArgArgGlySerGlyTyrHisSerAr 1823  
 Qy 975 GTATGCTCAGAGTCTACTACTATTGGAAGAAATTCGATGATGACCGCGTCTCGCTCG 1034  
 Db 1823 gSerProAlaArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgSerAr 1843  
 Qy 1035 TCTTGACGTAGTGTGAAGCGTCCGAGGAAGAGCATTC- - - GCAGACTGGATCCG 1091  
 Db 1843 gThrProThrSerArgLysArgSerArgSerArgThrSerProAlaProPlySerAr 1863  
 Qy 1092 TAAGAAGCGCAAGAGTG- - - - - CTGCTATGCCGATGATTTCTCTTC 1133  
 Db 1863 gSerArg- SerArgAlaSerProAlaThrHisArgArgSerArgSerArgThrProLeuI 1883  
 Qy 1134 TCTCGAAAGGCTTATAGGAAGGAGCCAGCCACCGTGGATGACTTATTTGAGCGA 1193  
 Db 1883 leSerArgArgSerArgSerArgThrSerProValSerArgArgSerArgSerAr 1903  
 Qy 1194 GA- - - - - CGCTCTTCGTTGTTACCGAGGTGGTTTCG 1223  
 Db 1903 rgThrSerValThrArgArgArgSerArgSerArgAlaSerProValSerArgArgAr 1923  
 Qy 1224 TTTTGCACAGTTTCCCAACGATTTGGCTTACAAATCCTGATGCTCATGCCGCTATCCTCAA 1283  
 Db 1923 erArgSerArgThrPro- - - - - ProValThrArgA 1933  
 Qy 1284 ATCGCTTGACGACAAAGTACAAAGACTACCTCCCTCGCTCGACCGTAAAGTGTGCGCGC 1343  
 Db 1933 rgArgSerArgSerArgThrProThrArgArgArgSerArgSerArg- - - ThrProp 1952  
 Qy 1344 CATGCTCGATATTGACGCGCGGTATCCCTGCGC- - - - - 1378  
 Db 1952 roValThrArgArgSerArgSerArgThrProProValThrArgArgSerArgAr 1972  
 Qy 1379 - - - - - ACAAGTCCCGCATATATTCAAGAATGTATCGACAAGAAATTCAAAGCGCACACAA 1436  
 Db 1972 erArgThrSerProIleThrArgArgSerArgSerArgThrSerProValThrArgA 1992  
 Qy 1437 GA- - - - - AGTATCGACACTTCGTATTTCGACAAAGAGTGTGTTCTTATACGCAAGATT 1490  
 Db 1992 rgArgSerArgSerArgThrSerProValThrArg- - - - - ArgAr 2006  
 Qy 1491 CCATGCCATGCTCAAGTCCATCGCAAGG- - - - - AAAAGTTTGGCAAGC 1535  
 Db 2006 erArgSerArgThrSerProValThrArgArgSerArgSerArgThrProAlaI 2026  
 Qy 1536 TATCGAAGAAAGATCCGCGCAGTAGAGC- - - - - TTTCCAAGAGCGTAATAGCTGCTGCTCG 1589  
 Db 2026 leArgArgArgSerArgSerArgThrProLeuLeuProArg- - - - - 2039  
 Qy 1590 CGCTATTTCAGGCGCATCGGATGGCCATGCTATGCCATTCGAGAGGCGCAAGCGTCTTTT 1649  
 Db 2040 - - - - - LysArgSerArgSerArgSerProLeuAlaIleArgArgSerArgSerAr 2057  
 Qy 1650 CTTTTCGCGTTTGGTGTAGATGTACCCCGGAGTCTCTGCGGAGCGATGCCAATTCAC 1709  
 Db 2057 rgThrProArgThrAlaArgGlyLys- - - - - ArgSerLeuThrArgSerP 2072  
 Qy 1710 CATCGGTATGAGCTACGCTCCATCAAGGGATATGAACCGCAGGCGTGCCTGCTACAA 1769  
 Db 2072 ro- - - - - ProAla- - - - - IleArgArgArgSerAlaSerGlySers 2084  
 Qy 1770 CTATCATCAGCACAGCAAGCGGCTATTGGAGAACAGGATCCTTAAGACCGATGACTTGC 1829  
 Db 2084 erSerAspArgSer- - - - - ArgSerAlaThrProp 2094  
 Qy 1830 CGTACAGGAGAAATATCTCGACCTTCCCGCACCAAAAATATCGTCCGTATGCGCGAGAA 1889  
 Db 2094 roAlaThrArgAsnHisSerGlySerArgThrPro- - - - - ProVala 2108  
 Qy 1890 CGGTACGCTCCATATATCGCTTTCTCTATCGAAACAGCATCATCAGGCGGTAACTCCGCTAG 1949

```
Db 2108 laLeuAnSerSerArgMetSerCysPheSerArgProSerMetSerProThrProLeuA 2128
QY 1950 CCCGTTATTCGATAAAGACGGCGTCTGATCGGTCTTTCGATCGCAACATGGGAAGC 2009
Db 2128 sp-----ArgCysArgSerProGlyMetL 2136
QY 2010 TATGATGGTGACATCGAGTTCGAAACCCGATCTCGACGCAATCAGCGTGGGA----- 2063
Db 2136 euGluProLeuGlySerSerArgThrProMetSerValLeuGln-GlnAlaGlyGlySer 2155
QY 2064 -----CATCCGCTACGTTCTTCATCATGATTCACAAATGGGGTCAAGTCCCGCTCT 2114
Db 2156 MetMetAspGlyProGlyProArgIleProAspHisGlnArgThrSerValProGluAsn 2175
QY 2115 CATCAAGAGCTGAAGTTGATCTA 2138
Db 2176 HisAlaGlnSerArgIleAlaLeu 2183

RESULT 11
Q9C105 PRELIMINARY; PRT; 1236 AA.
AC O9C105;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative glucoamylase I (alpha-1,4-glucan glucosidase), extracellular
DE starch-degrading enzyme, by similarity to S. cerevisiae STAl, contains
DE chitinase family signature.
GN SPAPBLE7.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Wood V., Rajandream M.A., Barrell B.G., Seeger K., Harris D.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; AL590805; CAC36921.1; -
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001917; NHTransf_2.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1236 AA; 123387 MW; 5A2D33A30B87CDD8 CRC64;

Alignment Scores:
Pred. No.: 0.0455 Length: 1236
Score: 154.50 Matches: 145
Percent Similarity: 35.92% Conservative: 110
Best Local Similarity: 20.42% Mismatches: 258
Query Match: 4.04% Indels: 198
DB: 3 Gaps: 24

US-10-008-355-1 (1-2139) x Q9C105 (1-1236)
QY 166 AGTTTCGACAAAGCGCTCATTCGCAATGCCCGTGTATCTTCGTTGGCGGATGACCGGT 225
Db 215 SerPheAspPheLeuSerIleHisThrPheAsnSerSerThrGlyGlyGlyCysSerGly 234
QY 226 -----ATCACAGTGTCCGATCAGGCGCTCATCTTTACC 258
Db 235 SerArgAsnSerThrPheAspAlaTrpValGluTyrAlaGluAspSerAlaTyrAsnThr 254
QY 259 AACCACTCGGATCGGTCGCTATCCAGACGCAAGCAGCGGTGATCAGGACTATCTG 318
Db 255 AsnThrSerLeuPheTyrGly-----ValValGlyHisGlnAsnGly 268
QY 319 CGCGATGGTTTCGTTCT---CGCACGATGGGTGAGGAGCTT----- 357
```

```
Db 269 SerAsnGlyPheIleSerProLysAsnLeuThrArgAspLeuLeuAsnTyrLysAlaAsn 288
QY 358 -----CCGATTCGGGGTCTTTCC----- 375
Db 289 SerThrLeuPheGlyGlyValThrIleTrpAspThrSerLeuAlaAlaMetSerTyrAsp 308
QY 376 -----GTGAAGTATCTGCGCAAGATCGTGAAGGTAAACGACAAGGTA 417
Db 309 AsnSerSerGluThrPheValGluAlaIleHisLysIleLeuAspThrLysSerLys--- 327
QY 418 GAAGGACAGCTCAAGGTATCACTGACGAGATGGAGCGTCTCGCAAGACTCAGAGGTA 477
Db 328 HisSerSerLysSerHisAspSerSerGlnGlyLeuGluSerThrSerSerIle 347
QY 478 TGCCAAGACTGGCCAAAAAAGAAATGCACAGAGAACCAACTCTGCATCGTAGAGCT 537
Db 348 -AlaLeuAsn---ProThrSerSerIleSerSerThrSerSerSerSerThrSerSe 366
QY 538 TTCTATTCCAAACGAATACTCTCTCATCTGCTACGATGATTTCAGGAGCGTTCTGATG 597
Db 366 rAlaIleSerThrIleSerGlnAspHisThrLysThrValThrSer----- 381
QY 598 GTATTTCCTCTCCAGCTCTGTAGTAAGTTCGGAGCGGATACGGACAACATGGATGCG 657
Db 382 -----ValSerAspGluProThrThrIleThrAlaSerG 393
QY 658 CCGCGTCAACGGCGGACTTCAGCGTATTCCGCGTGTATCCGGTGGCGGACAAACGGCGG 717
Db 393 y---AlaThrSerValThrThrThrThrLysThrAspPheAspThrValThrThrI 412
QY 718 GCCGATACACAGGAGCAATAAACCCCT-----ATAAGCCCTT 756
Db 412 evalSerThrSerThrLeuIleSerAlaSerAspSerThrSerIleIleValSerSerTy 432
QY 757 TACTTCGCTGCGGTATCCATCCAGGCTACAGGCTACAGGCTACGATGCGCATGACCATCG 816
Db 432 rValSerThrValThrGlnProAlaSerThrArgValGlnThrThrValSerSerI 452
QY 817 TTCGGGCGGACGTACGGATCGCTACCTCCTCTTGGGGTGGGAAGATCGTATCGAAAC 876
Db 452 eserThrSerVally-----GlnProTh 460
QY 877 GAGAACATCTCTGATCGAAGTTCGCGGTATCAAGCAAGCATGTGGAAGAAAGCCATG 936
Db 460 rAlaSerValAlaSerSerValSerValProSerSerSerValGln----- 477
QY 937 AGCGCAGATCAGGCTACCCGTATCAATATGCCAGCAAGTATGCTCAGAGTCTAATAT 996
Db 478 -----ProGlnSerSerThrProIleSer----- 485
QY 997 TGAAGAATTCGATCGGTATGAACCGCGTCTCGCTCTTTCGAGTGATAGTTCGTAAAG 1056
Db 486 -----Se 486
QY 1057 CGTCCCGGAGAAAGAGCATTCGCAGACTGGATCCCTAAGAACGGCAAGAGTCTGTCTAT 1116
Db 486 rSerSerSerAlaSerSerProGlnSerThrLeuSerThrSerSerGluValValSerG 506
QY 1117 GGCGATGATTGCTCTCTCTCGAAAGCGCTTATAGGAAGGAGCCAGCCCAACCGTGAG 1176
Db 506 uValSerSerThrLeuLeuSer----- 513
QY 1177 ATGACTTATTTGAGCGGACACGCTCTTCGGTGGTACCGAGGTGGTTCGTTTTCACAGTTT 1236
Db 514 -----GlySerSerAlaIlePro----- 519
QY 1237 GCCAACGCTTGGCTACAAATCCCTGATCGCGGTATCTCCTCAAAATCGCTTGACGAC 1296
Db 519 ----- 519
QY 1297 AAGTACAAGACTACCTCCCTCCCTCGCTCGACCGTAAGGTGCTGCCCGCCCATGCTCATATT 1356
Db 1356 ----- 1356
```

```
Db 520 -SerThrSerSerSerThrProSerSerSerIleLeuSerSerProMetThrSerValle 539
QY 1357 GTACGGCGCGTATCCCTCCGACAGCTCCCGCGATA---TATTCAAGAATGTAATCGAC 1413
Db 539 uSerSerSerIleProThrSerSerSerSerAspPheSerSerIleThrTh 559
QY 1414 AGAATTCAAAGGCGACGAGAGAGATGTCAGACTTCGTATTCGACAGAGTGTGGTT 1473
Db 559 rIleSerSerGlyIleSerSerSerIleProSerThrPheSerSerValSerSerI 579
QY 1474 CCTTATAGGACAGAGTTCATCCATCGCTCAAG-----TCCATGGA 1514
Db 579 eUeSerSerSerThrSerSerSerThrSerIleSerSerSerSerTh 599
QY 1515 CAAGGAAAAGTTTGCAAGGCTATCGAGAAAGATCCGGCAGTAGAGCTTTCCAAGAGCGT 1574
Db 599 rSerSerThrPheSerSerAlaSerThrSerPro-----SerSe 613
QY 1575 ATAGCTGCTCGCGGTATTCAGGCCGATCGGATGGCCAATGCTATGCCATTGAGAA 1634
Db 613 rIleSerSerSerIleSerSerSerSerThrIleLeuSerSer-ProThrProSerThrs 633
QY 1635 GGGCAAGCGCTTTCTTTGCGGTTCGTCGATGTACCCCGGACGTCTCCCGAG 1694
Db 633 eSerLeuMetIleSerSerSerIleIleSerGlySerSerSerIleLeuSerSers 653
QY 1695 CGATGCCAACTTCACATCGGTATGAGTACGGCTCCATCAAGGGATATGAACCCGAGGA 1754
Db 653 eIleSerThrIlePro-----IleSerSerSerLeuSerThrThrYrSerSerV 670
QY 1755 CGGTGCTGGTACACTATCATACAGACGGAAGCGGTATTGGAGACGAGTCCCTAA 1814
Db 670 aIleProSerSerSerThrLeuValSerSer-----SerSerSerLeuI 685
QY 1815 GAGCGATGATGTTGCGGTACAGAGAATCTCTGACCTTCCGACACCAAAACTATGG 1874
Db 685 leValSerSerPro-----ValAlaSerSerSerSerProIleProSers 702
QY 1875 TCGCTATCGGAGAACGGTCACTCATCTTCCTTATCGAACAACGACATCACGGG 1934
Db 702 eSerSerLeuValSerThrTyrSerAlaSerLeuSerAsnIleThrHisSerSerLeuS 722
QY 1935 CGGTAACTCCGTCACCGCGTATTG---ATAAGAACGGCGCTGATCGGCTTGTCTTT 1991
Db 722 eLeuThrAlaMetSerSerSerSerAlaIleProThrSerValAsnSerSerThrLeu- 741
QY 1992 CGATGCCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGAAC-----CCGA 2039
Db 742 -----IleThrAlaSerSerSerAsnThrLeuLeuSerSerI 754
QY 2040 TCTGACGCGACAAATCAGCGTGACATCGGCTACGTTCTTCTCATGATTGACAAATGGG 2099
Db 754 leThrSerSerSerAlaIleValSerSerThrThrValSerAsnIleSerSerAsnLeuP 774
QY 2100 TCAGTGGCCCGCTCATCCCAAGAGC 2125
Db 774 roSerAlaThrAlaSerSerGinSer 782

RESULT 12
O8WWQ5 PRELIMINARY; PRT; 2448 AA.
AC O8WWQ5;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Mucin 5 (Fragment).
CN MUC5AC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=TRACHEA;
RX MEDLINE=21426417; PubMed=11535137;
RA Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
RT repetitive region.";
RL Biochem. J. 358:763-772(2001).
DR EMBL; AJ298317; CAC83674.1; -.
DR InterPro; IPR001209; Ribosomal_S14.
DR InterPro; IPR002919; TILCysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF01826; TIL; 2.
DR Pfam; PF00094; vwd; 3.
DR SMART; SM00214; vwd; 3.
DR SMART; SM00216; VMC; 3.
DR PROSITE; PS00527; RIBOSOMAL_S14; UNKNOWN_2.
FT NON_TER 2448
SQ SEQUENCE 2448 AA; 260970 MW; A5C1BD627844D952 CRC64;

Alignment Scores:
Pred. No.: 0.113 Length: 2448
Score: 151.00 Matches: 188
Percent Similarity: 35.64% Conservative: 100
Best Local Similarity: 23.27% Mismatches: 308
Query Match: 3.95% Indels: 213
DB: 4 Gaps: 41

US-10-008-355-1 (1-2139) x Q8WWQ5 (1-2448)
QY 46 TTGGGTGCTTACGGGTAGCCAAAGCCGACAAAGCATGTGGCTCCTCAACGAA----- 99
Db 1637 LeuGlyGlnAspValIleCysSerHisThrGluGlyLeuLeuAsnLysAsnGln 1656
QY 100 -----CTCAATCAGGAGATCTGGAATCGAATCGGTGAGTCGCGCTTTACG 144
Db 1657 LeuProIleCysTyrAsnThrGlu-IleArgIleGlnCysCysGluThrValAsnVa 1676
QY 145 CTCCGCTTGGATTCGC-----TCTACAGTTTCGACAAAGCCGTCCTATGCCAATGCC 195
Db 1676 lCysArgAspIleThrArgLeuProLysThrValAlaThrThrArgProThrProHisPr 1696
QY 196 GTGGTTATCTTCGGTGGCGGATGACCGGTATCAGGTGTCGCGATCAGGGCCTGATCTTT 255
Db 1696 oThrGlyAlaGlnThrGlnThrThrPheThrHisMetProSerAlaSerThrGluGl 1716
QY 256 ACCAACCCACC-----ACTCGGATACGGTCTATCCAGAGCCAAAGC 297
Db 1716 nProThrAlaThrSerArgGlyGlyProThrAlaThrSerValThrGlnGlyThrHisTh 1736
QY 298 ACGGTGGATCAGGACTATCTGCGCGATGTTTCGTTCTCGCACGATGGGTGAG-----GA 353
Db 1736 rThrLeuValThrArgAsnCysHisProArgCysThrTrpThrLysTrpPheAspValAs 1756
QY 354 GCTTCCGATTCGGGTCTT-----TCCGTGAAGTATCTGCCGACAGATCGTGAAGGT 404
Db 1756 pPheProSerProGlyProHisGlyAspLysGluThrTyrAsnAsnIleleArgSe 1776
QY 405 AACGGACAAAGTAGAAGGACAGCTCAAGGTTATCACTACGAGATGGAGCGTCTCGCAA 464
Db 1776 rGlyGluLysIleCysArgArgProGluGluIleThrArgValGlnCysArgAlaLysSe 1796
QY 465 AGCTCAGGAGGTATGCCAAGAA-----CTGGCCAAAGAAAGAA--- 501
Db 1796 rHisProGluValSerIleGluHisLeuGlyGlnValValGlnCysSerArgGluGluGl 1816
QY 502 -----AATGCAGACGACGAAC-----CAACTCTGCATCGTAGACCTTT 539
Db 1816 yLeuValCysArgAsnGlnAspGlnGlnGlyProPheLysMetCysLeu----- 1832
QY 540 CTATTCCAAACAGAAATACTTCTCTCATGCTACCATGTTATTCAAGGAGCTTCGTATGCT 599
Db 1833 -----AsnTyrGluValArgValLeuCysCysGluThrProArgGlyCysHisMetTh 1850
```

```
Qy 600 ATTGCTCTCCAGCTCTGTAGTAAGTTCGGAGCGCATACGACAACTCGATGTGCC 659
Db 1850 r---ThrProGlySerThrSerSerProAlaGluThrProSerThrThr 1869
Qy 660 CCGTCACAGCGCGACTTCACGCTATCCCGGTGTATCCCGTCCGACACCGCCGCG 719
Db 1869 erLysThrThrGluIleGlnAlaSerGlySerSerAlaProSerSerThrProGlyThrV 1889
Qy 720 CG-----AATACAGCAGGACAATAAACCCCTTATAAGCCCGTTTACTTCGCTCGCGTATC 773
Db 1889 alSerLeuSerThrAlaArgThrThrProAlaProGlyThrAlaThrSerVal----- 1906
Qy 774 CATCAAGGCTACAGGCTGACGAGTATGCCATGACCATCCGTTTCCGGCGAGTACGGA 833
Db 1907 -----LysLysThrPheSerThrProSerPro---ProProValPro----- 1919
Qy 834 TCGCTACCTCACTCTTCTGGGTGTGGAAGATCGTATCGATAACAGACAGACAATCTCGTAT 893
Db 1920 --AlaThrSer-----ThrSerSerMetSerThrThrAlaProG 1932
Qy 894 CGAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTAC 953
Db 1932 lyThrSerValValSerSerLysProThrProThrGluPro-----S 1946
Qy 954 CCGTATCAAAATATGC---CAGCAAGATGCTCAGAGTCTCAACTATTGGA----- 1001
Db 1946 erThrSerSer-CysLeuGlnGluLeuCysThrTrpThrGluTrpIleAspGlySerTyr 1965
Qy 1002 -----GAATTCGATCGGTATGAACCGCGTCTCGCTCGCTCT 1037
Db 1966 ProAlaProGlyIleAsnGlyGlyAspPheAspThrPheGln-----Asn 1980
Qy 1038 TGAGGTGAT-----AGTCGTAAAGCGTTCGCGAGGAAG 1070
Db 1981 LeuArgaspGluGlyTyrThrPheCysGluSerProArgSerValGlnCysArgAlaGlu 2000
Qy 1071 AGCATTCCAGACTGG---ATCCGTAGAACGGCAAGAGTCTGTC-----TATGGCA 1121
Db 2001 Ser-PheProAsnThrProLeuGlyArgLeuGlyGlnAspValIleCysSerHisThrGl 2020
Qy 1122 TGTATTGCTCTCTCGAAAAGGCTTATAAGAGGAGCCCAAGCCCAACCGTGGATG-- 1179
Db 2020 uGlyLeuIleCysLeuAsnLysAsnGlnLeuProProIleCysTyrAsnTyrGluIleAr 2040
Qy 1180 -ACTTATTGAGCAGACGCTCTTCGGTGTACCGAGTGTGTTTTCGACACAGTTTGC 1238
Db 2040 gIleGlnCysCysGluThrValAsnValCysArgAspIleThrArgProProLys----- 2058
Qy 1239 CAACGCATTGGCTACAAATCCTGATGCTATGCGCGGTATCCTCAATCGCTTGAGACA- 1297
Db 2059 -ThrValAlaThrThrArgProThrProHis-ProThrGlyAlaGlnThrGlnThrThp 2078
Qy 1298 -----AGTACAAGACTACTCTCCCTCGCTCGACCGCTAAGG----- 1333
Db 2078 heThrThrHisMetProSerAlaSerThrGluGlnProThrAlaThrSerArgGlyGlyp 2098
Qy 1334 -----TGCTGCC 1340
Db 2098 roThrAlaThrSerValThrGlnGlyThrHisThrThrProValThrArgAsnCysHisP 2118
Qy 1341 CGCCATGCTCG---ATATTGACCGCGCGGTATCCTCCGCGACAGCTCCCG----- 1390
Db 2118 roArgCysThrTrpThrThrPheAspValAspPheProSerProGlyProHisGlyG 2138
Qy 1391 -----ATATTATCAAGATGTAATCGA-----CAAGAAATTCAAAGCGCAC 1433
Db 2138 lyAspLysGluThrTyrAsnAsnIle-IleArgSerGlyGluLysIleCysArgArgPro 2157
Qy 1434 GAAGAAGTATGACACTTCGTTATTCGACCAAGAGTGTGTTCTTATAGCGACAAAGTTCCA 1493
Db 2158 GluGluIleThrArgLeuGln----- 2164
Qy 1494 TGCCATGCTCAAGTCCATGGACAA-----GGAAGAAAGTTTGCAAGGC 1535
```

```
Db 2165 CysArgAlaLysSerHisProGluValSerIleGluHisLeuGlyGlnValGln--- 2183
Qy 1536 TATCGAAGAAAGATCCGCGAGTAGAGCTTTCCAAAGAGCCTAATAGCTGCTGCGGCTAT 1595
Db 2184 -----CysSerArgGlu 2187
Qy 1596 TCAGGCCGATCGGATGGCCAATGCTATGCTATGAGAGGCGCAAGCGTCTTTTCTTTCG 1655
Db 2188 GluGly-----LeuValCysArgAsnGlnAspGlnGlnGly--ProPheL 2202
Qy 1656 CGTTTTCGCTGAGATGTACCCGGACGTCTCTGCGGAGCGATGCCA----- 1702
Db 2202 yMetCysLeuAsnIleGluValArgValLeuCysCysGluThrProLysGlyCysProV 2222
Qy 1703 --ACTTCCACCATGCTATGAGTACGCTCCATCAAGGATATGAACCGCAGGACGGTGC 1760
Db 2222 alThrSerThrProVal-----ThrAlaProSerThrProSerGlyArgAlaIleSerP 2240
Qy 1761 CTGTTACAACATATCATACGACGCAAGGGGTATTGG---AGAAAGCAGGATCTTAAGAG 1817
Db 2240 ro-----ThrGlnSerThrSerSerTyrPelnLysSerArgThrThrL 2255
Qy 1818 CGATGAGTTTCGCTACAGGAGATATCTCGACCTCTTCGACCAAAACTATGGTCG 1877
Db 2255 euValThr-----ThrSerThrThrSerThrProGlnThrSerThrT 2269
Qy 1878 CTATGCCGAGACGCTCAGCTCCATATCGTTCCTTCTCAACACACGACATCACGGCGG 1937
Db 2269 hrTyrAlaHisThrThrSerThrSerAlaProThrAlaArgThrThrSer----- 2286
Qy 1938 TAACCTCCGCTAGCCCGCTATTCGATAAGAGCGCGCTGATCGGCTCTTTCGATGG 1997
Db 2287 -----AlaProThrThrSerThrThrSerValProThrThr-----SerThrI 2301
Qy 1998 CACTGGGAAGTATGAGTGTGATCGATCGAGTTCGAACCCGATCTGACGCGCACAAATCAG 2057
Db 2301 leSerGlyPro-----LysThrThrProSerProValProThrThrSerThrSera 2319
Qy 2058 CGTGACATCGCTACGTTCTTCTTCATGATTGACAAATGGGTGAGTCCCGCTCAT 2117
Db 2319 laAlaThrSerThrIleSerAlaProThrThrSerThrThrSerValProGlyThrT 2339
Qy 2118 CCAAGAGC 2125
Db 2339 hrProSer 2341
RESULT 13
Q29071 PRELIMINARY; PRT; 528 AA.
AC Q29071;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Gastric mucin (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Sulina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=95275264; PubMed=775593;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
RA LaMont J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
RT mucin."
RL Biochem. J. 308:89-96(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=YORKSHIRE/CHESTER WHITE/HAMPSHIRE CROSS; TISSUE=GASTRIC MUCOSA;
RX MEDLINE=94102478; PubMed=7506218;
RA Turner B.S., Bhaskar K.R., Hadzopoulou-Cladaras M., Specian R.D.,
```





```
QY 1931 -----CGGCGGTAACCTCGGTAGCCCGGTAT 1957
Db 452 ValGlnProSerSerSerAlaProThrThrSerAlaThrSerValGlnProSer 471
QY 1958 TCATAGAAGCGGCTCTGATCGGCTTCTTTCGATGCGCACTGGGAAGCTATGAGTG 1917
Db 472 SerSerSerProProfileSerSerThrThrSerValGlnProSer-----487
QY 2018 GTGACATCGAGTTCGAACCGGATCTGAGCGGCACAATCAGCGTGGACATCCGCTACGCTC 2077
Db 488 ---SerSerSerSerProThrThrSerThrThrSerValGlnProSerSerSerGly 506
QY 2078 TCTTCATGATTGCAATGGGTCAGTGGCCCGCTCTCATCCAGAGC 2125
Db 507 SerAlaProThrThrSerAlaThrSerValGlnProSerSerSer 522

RESULT 14
Q8WZX1
ID Q8WZX1 PRELIMINARY; PRT; 1037 AA.
AC
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Related to DNA repair endonuclease rad2.
GN B7N14.040.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL66986; CAD21041.1; -.
DR InterPro; IPR000513; Exo_N1.
DR InterPro; IPR002950; Josephin.
DR InterPro; IPR003903; UIM.
DR InterPro; IPR001532; XPGC_Rad.
DR Pfam; PF02809; UIM; 2.
DR Pfam; PF00867; XPG_I; 1.
DR Pfam; PF00752; XPG_N; 1.
DR SMART; SM00484; XPGI; 1.
DR SMART; SM00485; XPGN; 1.
KW Endonuclease.
SQ SEQUENCE 1037 AA; 114115 MW; C3977877D9444239 CRC64;

Alignment Scores:
Pred. No.: 0.139 Length: 1037
Score: 147.00 Matches: 152
Percent Similarity: 31.96% Conservative: 80
Best Local Similarity: 20.94% Mismatches: 225
Query Match: 3.85% Indels: 269
DB: 3 Gaps: 38

US-10-008-355-1 (1-2139) x Q8WZX1 (1-1037)
QY 58 GGGGTAGCCAAAGCCGACAAAGGCG-----ATGTGG-----CTCCTCAACGAA 99
Db 231 GlyIleArgArgAlaAspGluAlaLeuGluArgTrpLysGluArgLeuThrArgGlu 250
QY 100 CTC-----AATCAGGAGAACTCGATCGAATCGGTGAGCTCGGCTTACGCTCCCG--- 150
Db 251 LeuArgThrAsnGluSerGlyPhePheArgThrArgHisLysAlaLeuAspIleProGlu 270
QY 151 -----TTGGATTGCTCTACATGTTCCGACAGCGGTCATGCGCATGCCGCGTG 198
Db 271 GlyPheProSerLeuGluValLeuArgTyrThrHisProValValSerLysSerGlu 290
```

```
QY 199 GTATATC-----TTCGTGCGGATGTACCGGTATACAGTGTCCGAT 240
Db 291 ThrValGlnArgLeuLysArgGlnPheProGly-----LysGlnAspValAspVal 307
QY 241 CAGGCGCTG---ATCTTTACCAACCACTCGGATACGGTGTCTTCCAGAGCCAAAGC 297
Db 308 GluGlyLeuArgValPheVal-----ArgGlu 316
QY 298 ACGGTGATCAGCACTATCTCGCGATGCTGTTTCTGTTCT-----CGCACGATGGGTGAG 351
Db 317 ThrPheAspTrpGluTyr---ArgThrGlyAlaValLysPheIleArgValLeuAla--- 334
QY 352 GAGCTTCGATTCGCGGTCTTTCGTTGAAGTATCGCGCAAGATCGTGAAGTAAACGAC 411
Db 335 -----ProSerLeuLeuValGlnPheLeu-----LeuGln 344
QY 412 AAGTAGAAGACAGCTCAAGGGTATCACTGACGAGATGGAGCTCTGCGCAAGACTCAG 471
Db 345 ArgSerGluGlyAsp-----AspAspThrAspAspLeuArgArgLysGlu 359
QY 472 GAGGTATGCCAAGAACTGGCCAAAGAAATGCACAGCAGAGAACCACTCTGCTGCTGTA 531
Db 360 -----LysGluGluAla-----AlaPheVal 366
QY 532 GAGCTTTCTTATTCACAACAGTAATCTCTCATCTCTACGATGTTATCAAGAGAGCTT 591
Db 367 LysAlaIleLysSerArgArgThrHisPheSerThr-----AspAlaThrProGluLeu 384
QY 592 CGTATGTATTTGCTCTCCAGCTCTGTAGGTAAAGTTCGGAGCGCATACGGACAACGTG 651
Db 385 ArgIleSerTyrValPro-----390
QY 652 ATGTGGCGCGCTCACACGGCGGACTTTCAGCGTATTCGCGGTATCCGGTCCGCGCAAC 711
Db 391 -----IleLysValValThrLeuAspLeuAspAla 400
QY 712 CGCGCGCGCAATACAGCAAGGACAATAACCCCTTAAGCCGCTTACTTCGCTCGCGCTA 771
Db 401 GluProGluGlu-----GluAspAsnGluGlyPheGlyArgValGlyLeuAlaLeuAsn 418
QY 772 TCCATGCAAGGCTACAGGCTGAGCTATGCCATGACCATCGGTTTCCCGGGGAGTACG 831
Db 419 SerAspAspGluPheAspAlaGluAspGluThr-----ProGlySerThr 433
QY 832 GAT-----834
Db 434 GlnGlyGlySerSerAlaLysProPheAspProPheLysProAspLeuAlaTrpIlePro 453
QY 835 CGCTACCTCTACTTCTTGGGGTGTGGAAGATCGTATCGAAACACGAGACAATCCTCGTATC 894
Db 454 GluThrValAlaLysLeuGlyValProLeuThrValGlu-----466
QY 895 GAAGTTCGCGGTATCAAGCAAGGCATCTGGAAGAACCCATGACGCGCAGATCAGGCTACC 954
Db 467 -----IleTrpGluGlu-----LysGlnArgAlaLys 475
QY 955 CGTATCAATATGCCAGCAAGTATGCTCAGAGTCTCACTACTTTCGAAGAATTCGATCGGT 1014
Db 476 GluLeuLysAlaAlaAsnLysGlyThrArgLysAlaGlnThrThrLysGlnThrGlyGly 495
QY 1015 ATGAACCGCGCTCTGCTGCTTTCAGCTGATAGGTCTGTAAGCGTCCGAGGAAAGACGA 1074
Db 496 MetProAlaGly-----Ala 500
QY 1075 TTCGCACTGATGATCCGTAAAGACGCGCAAGAGTCTGCTGTATGGCGATGTATGCTTCT 1134
Db 501 LeuGluLysTyrValLysValThrLysGlySerAlaThrAlaGlyThrValThrLysSer 520
QY 1135 CTCGAAAGGCTTATAGGAGGAGGACCAAGCCCAACCTGAGATGACTTATTTGAGCGAG 1194
Db 521 -ThrSerAspTyrLeuSerLeuAspSerSerProProArgSerSerHisIleAlaProPr 540
```



Db	811	LeuTyrLeuTyrTyrLeuGlnAsnGlyArgAspMetTyrValAspGlnGluLeuAspIle	830
Qy	604	GCTCCTCCAGCTCTCTAGTAAAGTTCGGAGCGATACGGACAACCTGGATGTGGCCCGT	663
Db	830	-----	830
Qy	664	CACACGGCGACTTCACGCTATATCCCGGTGTATCCGGTCCGCAGACACCAGCGCGCGCA	723
Db	831	-----	835
Qy	724	TACAGACAGGACAATAAACCTATAAGCCGTTTACTTTCGCTCGCTATCCATCATCAAGGC	763
Db	836	TyrAspValAspHis-----	845
Qy	784	TACAGGCTGACGACTATGCCATGACCATCGTTTCCCGGCAGTACGGATCGCTACCTC	843
Db	846	PheLeuLysAspAsp-----	856
Qy	844	ACTTCTGGGTGTGGAAGATCGTATCGAAGACGAGACAATCCTGTATCGGAAGTTCCG	903
Db	857	LeuThrArgSerAspLysAsnArgGlyLysSerAspAsnValProSerGluGluValVal	876
Qy	904	GGTATCAAGCAAGCATCTCGAAGGAGCCATGACCGCAGATCAGGCTACCCGTATCAAA	963
Db	877	LysLysMetLysAsnTyrTirpArgGlnLeuLeuAsnAlaLysLeuIleThrGlnArgLys	896
Qy	964	TATGCCACGAAGTATGCTCAGAGTGTAACTATTTCGGAAGATTCGATCGGTATCAACCGC	1023
Db	897	Phe-----	906
Qy	1024	GGTCTCGCTCGTCTTGACTGTATAGT-----	1074
Db	907	GlyLeuSerGluLeuAspLysAlaGlyPheIleLysArgGlnLeuValGluThrArgGln	926
Qy	1075	TTCCGACAGTGGATC-----	1104
Db	927	IleThrLysHisValAlaGlnIleLeuAspSerArgMetAsnThrLysTyrAspGluAsn	946
Qy	1104	-----	1104
Db	947	AspLysLeuIleArgGluValLysValIleThrLeuLysSerLysLeuValSerAspPhe	966
Qy	1104	-----	1104
Db	967	ArgLysAspPheGlnPheTyrLysValArgGluIleAsnAsnTyrHisHisAlaHisAsp	986
Qy	1105	-----	1143
Db	987	AlaTyrLeuAsnAlaValValGlyThrAlaLeuIleLysLysTyrProLysLeuGluSer	1006
Qy	1144	GCTATTAAAGGAAGGA-----	1170
Db	1007	GluPheValTyrGlyAspTyrLysValTyrAspValArgLysMetIleAlaLysSerGlu	1026
Qy	1171	CGTGAGATGACTATTGTCAGCAGACGCTCTTCGTGTGTCACCGAGTGGTGTGTTTGC	1230
Db	1027	GlnGluIleGlyLysAlaThrAlaLysTyrPhePheTyrSerAsnIleMetAsnPhePhe	1046
Qy	1231	CAGTTTCCCAACCCATGGCTACAATCCTGATGCTCATCGCGGTATCCTCAATCG---	1287
Db	1047	LysThrGluIleThrLeuAlaAsnGlyGluIleArgLysArgProLeuIleGluThrAsn	1066
Qy	1288	-----	1329
Db	1067	GlyGluThrGlyGluIleValThrAspLysGlyArgAspPhe-----	1084
Qy	1330	AAGTGTGTCGCC---GCCATGCTCGATATGTACGCCGCGCT-----	1368
Db	1085	LysValLeuSerMetProGlnValAsnIleValLysLysThrGluValGlnThrGlyGly	1104
Qy	1369	-----	1401

Db	1105	PheSerLysGluSerIleLeuProLysArgAsnSerAspLysLeu-----	1119
Qy	1402	AATGTAATGACAGAANAATTCAAAGCGCACAGAAGATATGCAGACTTCGTATTCGAC	1461
Db	1120	-----IleAlaArgLysLysAspTrpAspProLysLysTyrGlyGly-----PheAsp	1135
Qy	1462	AAGAGTCGTGTTCTTATACGCACAAGTTCATGCC-----	1497
Db	1136	SerProThrValAlaIleTyrSerValLeuValAlaAlaLysValGluLysGlyLysSerLys	1155
Qy	1498	ATGCTCAAGTCC-----	1530
Db	1156	LysLeuLysSerValLysGluLeuLeuGlyIleThrIleMetGluArgSerSerPheGlu	1175
Qy	1531	-----	1554
Db	1176	LysAsnProIleAspPheLeuGluAlaLysGlyTyrLysGluValLysLysAspLeuIle	1195
Qy	1555	GTAGAGCTTTCCAAGACGCTAATAGCTGCTGCTCGCTATTACGCCGATGCCATGCC	1614
Db	1196	IleLysLeuProLys-----	1201
Qy	1615	AATGCCATGCGCATTCGAGAAGGGCAGCGCTTCTTTTCGGGTTTCGCGTGAGATGTAC	1674
Db	1202	SerLeuPheGluLeuGluAsnGlyArgLysArgMetLeuAlaSerAlaGlyGluLeuGln	1221
Qy	1675	CCCGGAGCT-----GCTCTGCCGAGCGAT---GCCAACTTCACCATCGTATGAGC---	1722
Db	1222	LysGlyAsnGluLeuAlaLeuProSerLysTyrValAsnPheLeuTyrLeuAlaSerHis	1241
Qy	1723	TACGGCTCCATCAGGCGATATGAACCGCAGGACGGT-----CCCTGGTGACAC	1770
Db	1242	TyrGluLysLeuLysGly---SerProGluAspAsnGluGlnLysGlnLeuPheValGlu	1260
Qy	1771	TATCATACGACAGCAAGCGCGATTGGAGAGCAGGATCCTCAAGCAGGATGAGTTCGCC	1830
Db	1261	GlnHis-----LysHisTyrLeuAspGluIleLeuGluGlnIleSerGluPheSer	1277
Qy	1831	GTACAG-----	1836
Db	1278	LysArgValIleLeuAlaAspAlaAsnLeuAspLysValLeuSerAlaTyrAsnLysHis	1297
Qy	1837	-----GAGATATCTCCGACCTCTCCCGCACCAAAAC	1869
Db	1298	ArgAsnLysProIleArgGluGlnAlaGluAsnIleIleHisLeuPheThrLeuThrAsn	1317
Qy	1870	TATGGT-----	1884
Db	1318	LeuGlyAlaProAlaAlaPheLysTyrPheAspThrThrIleAspArgLysArgTyrThr	1337
Qy	1885	GAGAACGGTCAGCTCCATATCGGTTCTTCCTATGCACACAGCAGATCCAGGGC	1935
Db	1338	SerThrLysGluLeuLeuAspAlaThrLeuIleHisGlnSerIleThrGlnIle	1354

Search completed: May 16, 2003, 12:35:34  
Job time : 221.5 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 04:53:13 ; Search time 3810 Seconds  
(without alignments)  
16338.819 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaaatgaattaaaaag.....aagagctgaattgatctaa 2139

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*

2: gb\_hgt:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vi:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vi:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rod:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_sy:\*

39: em\_htgo\_hum:\*

40: em\_htgo\_mus:\*

41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	1.0	5898	8	AY038360	Nectria h
2	21	1.0	165743	2	AL691459	Homo sapi
3	21	1.0	166143	9	AC013447	Homo sapi
4	21	1.0	200368	9	AC096541	Homo sapi
c 5	20	0.9	4309	2	AC017213	Drosophil
c 6	20	0.9	10689	1	AE004008	Xylella f
c 7	20	0.9	86519	9	AP001860	Homo sapi
c 8	20	0.9	129240	9	AC083826	Homo sapi
c 9	20	0.9	131813	2	AC011067	Drosophil
c 10	20	0.9	136241	2	AC120223	Rattus no
c 11	20	0.9	150505	2	AC095129	Rattus no
c 12	20	0.9	152120	9	AL391294	Human DNA
c 13	20	0.9	160279	2	AC019024	Homo sapi
c 14	20	0.9	160282	3	AC009253	Drosophil
c 15	20	0.9	163706	2	AC098197	Rattus no
c 16	20	0.9	185837	9	AC005552	Homo sapi
c 17	20	0.9	187700	2	AC079735	Mus muscu
c 18	20	0.9	169894	2	AC125392	Pan trogl
c 19	20	0.9	169891	2	AC104241	Homo sapi
c 20	20	0.9	171225	2	AC128439	Rattus no
c 21	20	0.9	172715	10	CNS07EGP	AL591826 BAC 13C18
c 22	20	0.9	177024	2	AC102780	Mus muscu
c 23	20	0.9	177262	2	AC123979	Papio cyn
c 24	20	0.9	183238	2	AC053525	Homo sapi
c 25	20	0.9	183915	2	AC073967	Homo sapi
c 26	20	0.9	186640	2	AC129152	Rattus no
c 27	20	0.9	187681	2	AC080124	Homo sapi
c 28	20	0.9	193134	2	AC016702	Homo sapi
c 29	20	0.9	195538	9	AC092733	Homo sapi
c 30	20	0.9	196423	2	AC116933	Papio cyn
c 31	20	0.9	210636	9	AC006443	Homo sapi
c 32	20	0.9	220848	2	AC091579	Homo sapi
c 33	20	0.9	223078	9	AL512504	Human DNA
c 34	20	0.9	228944	2	AL844530	Mus muscu
c 35	20	0.9	272605	3	AE003669	Drosophil
c 36	20	0.9	274875	2	AC127582	Mus muscu
c 37	20	0.9	297876	2	AC124538	Mus muscu
c 38	19	0.9	148	5	AY044428	Oncorhync
c 39	19	0.9	218	5	AF104614	Salmo sal
c 40	19	0.9	237	5	AF104589	Oncorhync
c 41	19	0.9	237	5	AF104590	Oncorhync
c 42	19	0.9	237	5	AF104607	Salmo sal
c 43	19	0.9	240	5	AF104608	Salmo sal
c 44	19	0.9	240	5	AF104610	Salmo sal
c 45	19	0.9	240	5	U80294	Oncorhynch

ALIGNMENTS

RESULT 1  
AY038360 5898 bp DNA linear PLN 09-AUG-2002  
LOCUS Nectria haematococca mpVI retrotransposon Nht2, complete sequence.  
DEFINITION  
ACCESSION AY038360  
VERSION AY038360.1 GI:14700031  
KEYWORDS  
SOURCE Nectria haematococca mpVI.  
ORGANISM Nectria haematococca mpVI  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreales; Nectriaceae; Nectria.  
1 (bases 1 to 5898)  
Shiflett,A.M., Enkerli,J. and Covert,S.F.  
REFERENCE Nht2, a copia LTR retrotransposon from a conditionally dispensable  
AUTHORS chromosome in Nectria haematococca  
TITLE

```

JOURNAL Curr. Genet. 41 (2), 99-106 (2002)
MEDLINE 22067396
PUBMED 12073091
REFERENCE 2 (bases 1 to 5898)
AUTHORS Shiflett,A.M., Enkerli,J. and Covert,S.F.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2001) Forest Resources, University of Georgia,
Brooks Drive, Athens, GA 30602-2152, USA
FEATURES
Source
Location/Qualifiers
1..5898
/organism="Nectria haematococca mpVI"
/isolate="156-30-6"
/db_xref="taxon:70791"
/chromosome="MAK1"
/note="mpVI"
repeat_region 1..5898
/note="degenerate copia-like retrotransposon"
repeat_region 1..5
/transposon="Nht2"
1..5
/note="5' duplicated target site"
/rpt_type=direct
6..233
/misc_feature <3984..>5537
/note="similar to reverse transcriptase and RNase H
domains"
LTR 5666..5898
repeat_region 5894..5898
/note="3' duplicated target site"
/rpt_type=direct
BASE COUNT 1766 a 1483 c 1265 g 1383 t 1 others
ORIGIN
Query Match 1.0%; Score 21; DB 8; Length 5898;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1137 CGAAAGGCTTATAAGGAGG 1157 165743 bp DNA linear HTG 17-AUG-2002
|||||
DB 1864 CGAAAGGCTTATAAGGAGG 1884

RESULT 2
AL691459 165743 bp DNA linear HTG 17-AUG-2002
LOCUS Homo sapiens chromosome 1 clone RP11-543D5, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL691459
VERSION AL691459.24 GI:22415898
KEYWORDS HTG; HTGS-PHASE2; HTGS-ACTIVEPIN; HTGS-DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165743)
Direct Submission
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:22265419.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA543D5
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator; 15% of reads
Chemistry: Dye-terminator Big Dye; 84% of reads
Consensus quality: 165711 bases at least Q40

Consensus quality: 165743 bases at least Q30
Consensus quality: 165743 bases at least Q20
Insert size: 165743; sum-of-contigs
Insert size: 179024; 7.4% error; agarose-fp
Quality coverage: 11.08x in Q20 bases; sum-of-contigs Quality
coverage: 10.63x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source
Location/Qualifiers
1..165743
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-543D5"
/clone_lib="RPCI-11.2"
1..165743
/note="assembly_fragment:02071
clone_end:17
vector_side:left"
BASE COUNT 41605 a 39965 c 40507 g 43666 t
ORIGIN
Query Match 1.0%; Score 21; DB 2; Length 165743;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 GCACGATGGTGAGGAGCTTC 358 166143 bp DNA linear PRI 09-MAY-2001
|||||
DB 134330 GCACGATGGTGAGGAGCTTC 134350

RESULT 3
AC013447 166143 bp DNA linear PRI 09-MAY-2001
LOCUS Homo sapiens BAC clone RP11-543D5 from 1, complete sequence.
DEFINITION AC013447
ACCESSION AC013447
VERSION AC013447.4 GI:10334968
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166143)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
MEDLINE 9847074
PUBMED
REFERENCE 2 (bases 1 to 166143)
AUTHORS Joshua,C., Stoneking,T., Gregory,S. and Burkett,M.
TITLE The sequence of Homo sapiens BAC clone RP11-543D5
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 166143)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 166143)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 166143)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 166143)
REFERENCE

```

**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

Waterston, R.  
 Direct Submission  
 Submitted (09-MAY-2001) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Sep 28, 2000 this sequence version replaced gi:7631057.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0543D05  
 -----

**NOTICE:** This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBAC3.6

**NEIGHBORING SEQUENCE INFORMATION:**

The clone sequenced to the left is RP11-689L22. Actual start of this clone is at base position 1 of RP11-543D5; actual end is at base position 166143 of RP11-543D5.

The sequence fidelity of RP11-543D5 from base position 100749 to 101626 can not be guaranteed. This region is made up of a variable GT run where there are single M13 and PCR only regions.

**FEATURES**

Source	1. .166143 Location/Qualifiers
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="1"
	/clone="RP11-543D5"
	/clone_lib="RP11-11"
repeat_region	79..574 /rpt_family="L2"
repeat_region	2943..3362 /rpt_family="L2"
repeat_region	3930..4010 /rpt_family="MIR"
repeat_region	4107..4194 /rpt_family="MIR"
repeat_region	4410..4766 /rpt_family="L2"
repeat_region	4783..4857 /rpt_family="MIR"
repeat_region	5456..5581 /rpt_family="L2"
	5631..5737 /rpt_family="CRI"
	6897..6961 /rpt_family="L1"
	7348..7465 /rpt_family="MIR"
	7567..7703 /rpt_family="MIR"
	9706..9809 /rpt_family="CRI"
	10546..10648 /rpt_family="L2"
	10649..10945 /rpt_family="Alu"
	10946..11080 /rpt_family="L2"
	11081..11462 /rpt_family="MaLR"
	11463..11526 /rpt_family="L2"
	11774..11894 /rpt_family="MIR"
	12407..12616 /rpt_family="MIR"
	13849..13950 /rpt_family="L2"
	15240..15467 /rpt_family="MIR"
	15520..15634 /rpt_family="MIR"
	15812..15969 /rpt_family="L2"
	16021..16139 /rpt_family="L2"
	16457..16677 /rpt_family="MIR"
	17300..17592 /rpt_family="L1"
	17826..17960 /rpt_family="L1"
	18236..18379 /rpt_family="MIR"
	18565..18980 /rpt_family="ERV1"
	18993..19107 /rpt_family="MIR"
	19181..19473 /rpt_family="Alu"
	19532..19756 /rpt_family="L1"
	19757..19930 /rpt_family="Alu"
	19931..19996 /rpt_family="L1"
	20075..20213 /rpt_family="MIR"
	20328..20518 /rpt_family="L2"
	20638..20929 /rpt_family="MER1_type"
	20958..21229 /rpt_family="MaLR"
	21230..21277 /rpt_family="L2"
	21449..21587 /rpt_family="MaLR"
	21618..21668 /rpt_family="L2"
	22345..22655 /rpt_family="Alu"
	22821..23158 /rpt_family="L2"
	23309..23379



```

/rpt_family="MIR"
23705. .23875
/rpt_family="MaLR"
24042. .24152
/rpt_family="MaLR"
24440. .24732
/rpt_family="Alu"
24841. .25234
/rpt_family="MaLR"
25384. .25642
/rpt_family="MIR"
25694. .25846
/rpt_family="MIR"
26149. .26345
/rpt_family="MIR"
26440. .26539
/rpt_family="L2"

```

Query Match 1.0%; Score 21; DB 9; Length 166143;

Best Local Similarity 100.0%; Pred. No. 12; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 338 GCACGATGGTGAGGAGCTTC 358
      |||||
Db 134730 GCACGATGGTGAGGAGCTTC 134750

```

## RESULT 4

```

AC096541
LOCUS AC096541 200368 bp DNA linear PRI 31-MAY-2002
DEFINITION Homo sapiens chromosome 1 clone RP11-330M19, complete sequence.
ACCESSION AC096541 AL359084
VERSION AC096541.2 GI:21281550
KEYWORDS HTG.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 200368)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.
and Haugen, E.D.

```

## Direct Submission

```

JOURNAL Unpublished
2 (bases 1 to 200368)
Kaul, R.K., Olson, M.V., Raymond, C., Clendinning, J. and Haugen, E.D.
AUTHORS Direct Submission
TITLE Submitted (18-SEP-2001) Genome Center, University of Washington,
JOURNAL Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 200368)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimmachak, C., Phelps, K.A., Buckley, D., Kibukawa, M., Raymond, C.
and Haugen, E.D.

```

## Direct Submission

```

JOURNAL Submitted (31-MAY-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT On May 31, 2002 this sequence version replaced gi:15638689.

```

```

----- Genome Center
Center: University of Washington Genome Center
Center Code: UNGC
Web site: http://www.genome.washington.edu
Contact: uwgchgsu@washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP11-330M19 (sc0667)

```

```

----- Summary Statistics
Sequencing vector: plasmid; 38% of reads
Sequencing vector: plasmid; L08752; 62% of reads
Chemistry: Dye-terminator ET; 89% of reads
Chemistry: Dye-terminator Big Dye; 11% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 199860 bases at least Q40

```

```

Consensus quality: 200007 bases at least Q30
Consensus quality: 200368 bases at least Q20
Insert size: 200368; sum-of-contigs
Quality coverage: 9.5x in Q20 bases; sum-of-contigs

```

## Overlapping Sequences:

```

5': Mapping in progress
3': RP11-307017 AL451079

```

## Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

## Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HindIII			EcoRI			BglII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
859	862	8696	8787	2761	2748				
6382	6512	6	<800	2067	2051				
512	<800	169	<800	9556	9558				
449	<800	7794	7749	751	797				
5425	5304	1389	1365	1632	1611				
1893	1888	1144	1129	5843	5875				
538	<800	8353	8399	5875	5875				
567	<800	1209	1199	100	<800				
4932	4924	4077	4005	6408	6406				
3378	3413	1150	1129	7217	7267				
1519	1502	505	<800	108	<800				
1788	1776	3804	3782	4723	4657				
4138	4136	8058	8027	255	<800				
468	<800	975	964	708	<800				
709	<800	470	<800	549	<800				

```
----- 2417 2375 859 ----- 4278 4222 -----
----- 1016 1010 681 ----- 840 797 -----
----- 4965 4924 9699 ----- 16573 16486 -----
----- 1391 1362 1738 ----- 4181 4222 -----
----- 1828 1776 5846 ----- 344 <800 -----
----- 474 <800 7651 ----- 9127 9122 -----
----- 5588 5553 7 ----- 1212 1203 -----
----- 1009 1010 3687 ----- 4184 4222 -----
----- 8222 8258 1676 ----- 794 797 -----
----- 4161 4136 649 ----- 2549 2589 -----
----- 2143 2146 4518 ----- 4070 4033 -----
----- 13940 13696 991 ----- 20135 20286 -----
----- 7650 7663 13108 ----- 3615 3595 -----
----- 994 1010 59 ----- 1653 1611 -----
----- 2664 2660 3792 ----- 3390 3401 -----
----- 67 <800 1237 ----- 4250 4222 -----
----- 4553 4490 905 ----- 925 934 -----
----- 1617 1598 139 ----- 5490 5551 -----
----- 727 722 6393 ----- 197 <800 -----
----- 12486 12224 2138 ----- 3602 3595 -----
----- 2982 3036 1987 ----- 218 <800 -----
----- 5411 5304 823 ----- 9719 9558 -----
----- 3361 3413 1978 ----- 588 <800 -----
----- 7890 7901 2403 ----- 7908 8047 -----
----- 2581 2660 2790 ----- 1766 1766 -----
----- 5622 5553 13902 ----- 11220 10965 -----
----- 128 <800 2548 ----- 61 <800 -----
----- 232 <800 61 ----- 960 934 -----
----- 84 <800 2996 ----- 6639 6761 -----
----- 13376 13696 1777 ----- 2476 2589 -----
----- 8244 8258 1039 ----- 88 <800 -----
----- 1510 1502 12719 ----- 4639 4657 -----
----- 5987 6004 389 ----- 3630 3595 -----
----- 1609 1598 339 ----- 10219 10176 -----
----- 2327 2375 5609 ----- 3324 3401 -----
----- 13583 13696 1789 ----- 296 <800 -----

Query Match 1.0%; Score 21; DB 9; Length 200368;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 GCACGATGGGTGAGGAGCTTC 358
Db 12725 GCACGATGGGTGAGGAGCTTC 12745
|||||
|||||

RESULT 5
AC017213/c
LOCUS AC017213
DEFINITION Drosophila melanogaster, 4309 bp DNA linear HTG 09-DEC-1999
pieces.
ACCESSION AC017213
VERSION AC017213.1 GI:6553773
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
Adams, M. and Venter, J. C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10209899 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..4309
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 1270 a 846 c 867 g 1326 t
ORIGIN

Query Match 0.9%; Score 20; DB 2; Length 4309;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCAATGAAATTAAGTA 22
Db 2113 GCAATGAAATTAAGTA 2094
|||||

RESULT 6
AE004008/c
LOCUS AE004008
DEFINITION Xylella fastidiosa 9a5c, section 154 of the complete genome.
ACCESSION AE004008
VERSION AE004008.1 GI:9106961
KEYWORDS Xylella fastidiosa 9a5c.
SOURCE Xylella fastidiosa 9a5c.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.
REFERENCE 1 (bases 1 to 10689)
AUTHORS Simpson, A. J., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M.,
```

Alvarenga, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R., Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carter, H., Coutinho, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M., Coutinho, L.B., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H., Facincani, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohne, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P. and Marino, C.L.

**TITLE**  
The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis

**JOURNAL**  
**MEDLINE**

Nature 406 (6792), 151-157 (2000)

**PUBMED**  
20365717

**REFERENCE**  
10910347

**AUTHORS**  
Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M.C., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.S., Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carter, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M., Coutinho, L.B., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H., Facincani, A.P., Ferreira, A.J.S., Fraga, J.S., Franca, S.C., Franco, M.C., Frohne, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F., Lambais, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N., Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L., Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C., Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, M.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tshako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M., Meidanis, J. and Setubal, J.C.

**TITLE**  
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

**FEATURES**  
**source**

1. .10689  
/organism="Xylella fastidiosa 9a5c"  
/db\_xref="taxon.160492"  
/clone="9a5c"  
complement(119. .367)  
/gene="XF1871"  
complement(119. .367)  
/gene="XF1871"  
complement(119. .367)  
/notes="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

**gene**

**CDS**

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF84677.1"  
/db\_xref="GI:9106962"  
/translation="MTMTLKEDKTEIIDQRIQIAKLSETAKIQAEIHEVNAHTOKL MKETLKTESKWPVVGSSLLAAGATATLFLKLFH"  
complement(430. .618)  
/gene="XF1872"  
complement(430. .618)  
/gene="XF1872"  
complement(430. .618)  
/notes="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

**gene**

**CDS**

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84682.1"  
/db\_xref="GI:9106967"  
/translation="MFSWLMLRLNLIGOTVARGLVGAGLALVTTVPLIPLVTSALNL IYSKSGISADVNLNLIALMGEGEALSISGAMLTRALQSLHVGIVKATF"  
complement(3195. .4412)  
/gene="XF1877"  
complement(3195. .4412)

/product="hypothetical protein"  
/protein\_id="AAF84678.1"  
/db\_xref="GI:9106963"  
/translation="MMEERATAPQGGLAASDPQRTDAPKKTGLRKRKYPTWGSASI IVLRHVLTQRRKASP"  
941. .1288  
/gene="XF1873"  
941. .1288  
/gene="XF1873"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF84679.1"  
/db\_xref="GI:9106964"  
/translation="MSKNKIPNOILTGNAGFOFINGKLVTPERRTTEEMQLRWLSLT CTLARSWKMEERATAPQGGLAASDPQRTDAPKKTGLRKRHAANEVASTIQLRQ VLIQROOKRKASP"  
1351. .1662  
/gene="XF1874"  
1351. .1662  
/gene="XF1874"  
complement(1839. .2918)  
/gene="XF1875"  
complement(1839. .2918)  
/notes="similar to SP|P03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF84680.1"  
/db\_xref="GI:9106965"  
/translation="MTMTLKEDKTEIIDQRIQIAKLSETAKIQAEIHEVNTHTOEM NAHINEMNARTQEMNARTOKLIEETQVKESQWYPPVVASGLAAGATAATLFLKLF H"  
complement(1839. .2918)  
/gene="XF1875"  
complement(1839. .2918)  
/gene="XF1875"  
complement(1839. .2918)  
/notes="similar to SP|P03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1  
/transl\_table=11  
/product="phage-related protein"  
/protein\_id="AAF84681.1"  
/db\_xref="GI:9106966"  
/translation="MLNLVTGAPGNGKTLIYAVDWLIRQIEIDKSLVKSAGVPRSFYTD IEGFDVAVRLTGYVVSQAPEDWRTTPOGSIYVYDEAHMFAPGRSDDDPRVCDL DTHRHGSDLMFTVOTPKTHHELRLVGEHVLNRMGLTAGLVRMSAODDPYDI HQREKAEVEWKPDKRYALYASSTLHTVSHKFRIPKVKNSALSCVTCVIGLVFWH YTSFAHLSEAASSVAGASQASLURAPASLSRSSRLVSGMRTYAVLETESAPLSCGV SDRSCRCNTDGYQIDMSVVECRLLASPLFPNVYHAYVTSSSAAAASVSSPSSPS SSSVLSVSSVSSAPILGSSVPSH"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/gene="XF1876"  
complement(2918. .3202)  
/notes="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

```

/gene="XF1877"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF04683.1"
/db_xref="GI:9106968"
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF04683.1"
/db_xref="GI:9106968"
/translation="MCSRSLFVIFISALGLTLVIFPCLFAQNVPTVEWPTSVRY
SSVSDGVSTSAFVEARSTALVGRVYTPVDISASTGLSLAKVARRGMAFVNVY
MLSLINGAGWIDELTHEYMSGPALKIPIVGVVWVWQPCDGHFYSYTPQGLTAP
INAYQSLPQLVLTSSGFSVSSERWHLHGGGEYIHLKTDOSVDYSSGLPP
SVLDTDLGLVSDSTNAVLTDSQTGAVLITPEIVSALNKLRRSLDELEKASHAP
DQPSGGASPPSSSGTAWPFCFSCWASVYKTFIDWKDFLKLPLVPDVPYVDKL
PQATWSGGLGEGACPSPTPIPIEFSGYKTSVEISYQPFCDFAALMRPVVIVATILA
AYTAGGPRGVKNV"
/complement(4481..4702)
/gene="XF1878"
/complement(4481..4702)
/gene="XF1878"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84684.1"
/db_xref="GI:9106969"
/translation="MKSVIDRGIAFFSLTSSLVFAADAGASFDAAGAAVTALGATS
GAVALIGAALKLAPPAISVGMKWLKGAIFG"
/complement(4724..4924)
/gene="XF1879"
/complement(4724..4924)
/gene="XF1879"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84685.1"
/db_xref="GI:9106970"
/translation="MAQLMSLYCTQYDVEARTCSQAWVPVPSLLPPISYEDVRILL
PHVWCFVAVGHHFLFTVVRD"
/complement(4931..5236)

gene
Query Match 0.9%; Score 20; DB 1; Length 10689;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 646 AACTGGATGTGGCCGCGTCA 665
Db 9863 AACTGGATGTGGCCGCGTCA 9844

RESULT 7
AP001860/c
LOCUS 86519 bp DNA linear PRI 20-MAR-2001
DEFINITION Homo sapiens genomic DNA, chromosome 4q22-q24, clone:343A4,
complete sequence.
ACCESSION AP001860
VERSION AP001860.2 GI:13111481
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:343A4.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
Tsai,S.F.
Genome Sequencing of the Chromosome 4q Region Implicated in Human
Hepatocellular Carcinoma Pathogenesis 1
JOURNAL unpublished
REFERENCE 2 (bases 1 to 86519)
Tsai,S.F.
Direct Submission
TITLE

```

```

Submitted (24-APR-2000) Shih-Feng Tsai, National Yang-Ming
University, Institute of Genetics; 155 Li-Rong St. Section 2,
Peitou, Taipei, Taiwan 11221, Republic of China
(E-mail: ympetsai@ym.edu.tw, URL: http://genome.ym.edu.tw/,
Tel: 886-2-28267043; Fax: 886-2-28264930)
On Feb 22, 2001 this sequence version replaced gi:7649705.
Quality: the expected Phred/Phrap calculated error rate (per 10 kb)
is 0.0019; Estimated total number of errors is 0.0167.
FEATURES
Location/Qualifiers
source
1. 86519
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q22-q24"
/clone="343A4"
BASE COUNT 29441 a 16281 c 16491 g 24306 t
ORIGIN
Query Match 0.9%; Score 20; DB 9; Length 86519;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 TGAATTAATAAAGTATTCTT 27
Db 36681 TGAATTAATAAAGTATTCTT 36662

RESULT 8
AC083826/c
LOCUS 129240 bp DNA linear PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-289C17 from 4, complete sequence.
ACCESSION AC083826
VERSION AC083826.7 GI:14589704
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 129240)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
PUBMED 99063792
9847074
REFERENCE 2 (bases 1 to 129240)
VanBrunt,A. and Meyer,R.
The sequence of Homo sapiens BAC clone RP11-289C17
Unpublished
REFERENCE 3 (bases 1 to 129240)
Waterston,R.H.
Direct Submission
Submitted (02-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 129240)
Waterston,R.H.
Direct Submission
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 129240)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 3, 2001 this sequence version replaced gi:14269820.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0289c17

```

-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

#### VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-10L12. 2000 bp overlap: the clone sequenced to the right is AC022793. Actual start of this clone is at base position 156151 of RP11-10L12; actual end is at base position 129240 of RP11-289C17.

Data from AC08124 was used to finish this clone, AC083826. The region from 45337 to 45361 is only covered by data from AC08124. Polymorphisms exist between RP11-289C17 and RP11-10L12.

#### FEATURES

##### source

1. 129240  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-289C17"  
/clone\_l1b="RPCI-11"  
25. 309  
/rpt\_family="Alu"

##### repeat\_region

222. 662  
/note="match to EST AI791382 (NID:g5339098) oh01a11.y5"

##### misc\_feature

273. 756  
/note="match to EST AI732988 (NID:g5054101) oh01a11.x5"

##### repeat\_region

949. 1263  
/rpt\_family="Alu"

##### repeat\_region

1234. 1274  
/rpt\_family="(A)n"

##### repeat\_region

1310. 1593  
/rpt\_family="L1"

##### repeat\_region

1594. 1870  
/rpt\_family="Alu"

##### repeat\_region

1871. 2198  
/rpt\_family="L1"

##### misc\_feature

2223. 2226  
/note="similar to Homo sapiens EST AV650986 (NID:g9872000)"

##### misc\_feature

2223. 2226  
/note="similar to Homo sapiens EST BG291978 (NID:g13050347)"

##### repeat\_region

2471. 2626  
/rpt\_family="L1"

##### repeat\_region

repeat\_region  
2921. 2948  
/rpt\_family="AT-rich"  
3597. 3662  
/rpt\_family="L2"  
3814. 4312  
/rpt\_family="ERV"  
4361. 4495  
/rpt\_family="L1"  
4501. 4684  
/rpt\_family="ERV1"  
4644  
/note="similar to Homo sapiens EST BG386439 (NID:g13279885)"  
4681. 5216  
/note="similar to Mus musculus EST AA059627 (NID:g1553413)"  
mj74902.r1  
4681. 5087  
/note="similar to Mus musculus EST BF462055 (NID:g11531238)"  
4681. 5040  
/note="similar to Homo sapiens EST BF906415 (NID:g1297873)"  
4691. 4992  
/note="similar to Bos taurus EST BE484374 (NID:g9603907)"  
4697. 5115  
/note="similar to Homo sapiens EST BG386439 (NID:g13279885)"  
4701. 5382  
/note="similar to Homo sapiens EST BG576912 (NID:g13584565)"  
4724  
/note="similar to Homo sapiens EST BG827066 (NID:g14174653)"  
4728. 5272  
/note="similar to Homo sapiens EST BGL78667 (NID:g12685370)"  
4728. 5000  
/note="similar to Homo sapiens EST BE296600 (NID:g9180164)"  
4734. 5379  
/note="similar to Homo sapiens EST AUL33548 (NID:g10994087)"  
4734. 5275  
/note="similar to Homo sapiens EST BG757135 (NID:g14067788)"  
4734. 5179  
/note="similar to Homo sapiens EST BE018798 (NID:g8278841)"  
bb85b04.y1  
4734. 5095  
/note="similar to Homo sapiens EST BF527247 (NID:g11614610)"  
4743. 5542  
/note="similar to Mus musculus EST BG923425 (NID:g14303901)"  
4858. 5275  
/note="similar to Homo sapiens EST BG828340 (NID:g14175927)"  
4859. 5379  
/note="similar to Homo sapiens EST BG740954 (NID:g14051607)"  
4859. 5357  
/note="similar to Homo sapiens EST BG290663 (NID:g13047851)"  
4859. 5115  
/note="similar to Homo sapiens EST AW367838 (NID:g6872488)"  
4860. 5383  
/note="similar to Homo sapiens EST AW860531 (NID:g7956224)"  
4860. 5115  
/note="similar to Homo sapiens EST BE007807 (NID:g8268040)"

```

misc_feature 4865..5343
/notes="similar to Homo sapiens EST BE779572
(NID:g10200770)"
misc_feature 4870..5387
/notes="similar to Homo sapiens EST BG283401
(NID:g13033252)"
misc_feature 4874..5379
/notes="similar to Homo sapiens EST AW991862
(NID:g8251932)"
misc_feature 4875..5383

Query Match 0.98; Score 20; DB 9; Length 129240;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TGAATTAATAAAGTATTCCTT 27
|||||
Db 12472 TGAATTAATAAAGTATTCCTT 12453

.RESULT 9
AC011067
LOCUS AC011067 131813 bp DNA linear HTG 20-MAR-2000
DEFINITION Drosophila melanogaster chromosome 2 clone BACR09L20 (D1140)
RCR1-98.09.L.20 map 39A-39A strain Y; cn bw sp, *** SEQUENCING IN
PROGRESS ***, 130 unordered pieces.
AC011067
VERSION AC011067.4 GI:7264763
KEYWORDS HTG; HTGS, PHASE1.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 131813)
Cainiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,A.R., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Swirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 131813)
Cainiker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (01-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 20, 2000 this sequence version replaced gi:6984326.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 130 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
803: contig of 803 bp in length

```

```

804 883: gap of unknown length
884 1597: contig of 714 bp in length
1598 1677: gap of unknown length
1678 2236: contig of 559 bp in length
2237 2316: gap of unknown length
2317 2921: contig of 605 bp in length
2922 3001: gap of unknown length
3002 3884: contig of 803 bp in length
3885 4416: contig of 532 bp in length
4417 4496: gap of unknown length
4497 5056: contig of 560 bp in length
5057 5136: gap of unknown length
5137 5947: contig of 811 bp in length
5948 6027: gap of unknown length
6028 6767: contig of 740 bp in length
6768 6847: gap of unknown length
6848 7486: contig of 639 bp in length
7487 7566: gap of unknown length
7567 8210: contig of 644 bp in length
8211 8291: gap of unknown length
8291 9034: contig of 743 bp in length
9034 9113: gap of unknown length
9113 9770: contig of 657 bp in length
9770 9850: gap of unknown length
9851 10597: contig of 747 bp in length
10598 11128: contig of 451 bp in length
11129 12081: gap of unknown length
12082 12161: gap of unknown length
12162 13148: contig of 987 bp in length
13149 13228: gap of unknown length
13229 14305: contig of 1077 bp in length
14306 14385: gap of unknown length
14386 15384: contig of 999 bp in length
15385 16434: contig of 970 bp in length
16435 16514: gap of unknown length
16515 17933: contig of 1419 bp in length
17934 18013: gap of unknown length
18014 18888: contig of 875 bp in length
18889 19673: contig of 705 bp in length
19674 19753: gap of unknown length
19754 20616: contig of 783 bp in length
20617 21346: contig of 730 bp in length
21347 22580: contig of 1154 bp in length
22581 22660: gap of unknown length
22661 23567: contig of 907 bp in length
23568 24529: gap of unknown length
24529 24609: gap of unknown length
24610 25686: contig of 1077 bp in length
25687 25766: gap of unknown length
25767 27466: contig of 1700 bp in length
27467 27546: gap of unknown length
27547 29292: contig of 1746 bp in length
29293 29372: gap of unknown length
29373 30275: contig of 903 bp in length
30276 30355: gap of unknown length
30356 31407: contig of 1052 bp in length
31408 31487: gap of unknown length
31488 32729: contig of 1242 bp in length
32730 32809: gap of unknown length
32810 34227: contig of 1418 bp in length
34228 34307: gap of unknown length
34308 35359: contig of 1052 bp in length
35360 35433: gap of unknown length
35440 37214 37293: gap of unknown length

```

\* 37294 38442: contig of 1149 bp in length  
 \* 38443 38522: gap of unknown length  
 \* 39673 39673: contig of 1445 bp in length  
 \* 39968 40047: gap of unknown length  
 \* 40048 41551: contig of 1504 bp in length  
 \* 41552 41631: gap of unknown length  
 \* 41632 43684: contig of 2053 bp in length  
 \* 43685 43764: gap of unknown length  
 \* 43765 45714: contig of 1950 bp in length  
 \* 45715 45794: gap of unknown length  
 \* 45795 47443: contig of 1649 bp in length  
 \* 47444 47524: gap of unknown length  
 \* 47524 49494: contig of 1971 bp in length  
 \* 49495 49574: gap of unknown length  
 \* 49575 52237: contig of 2663 bp in length  
 \* 52238 53117: gap of unknown length  
 \* 53118 53865: contig of 1548 bp in length  
 \* 53866 53945: gap of unknown length  
 \* 53946 55202: contig of 1257 bp in length  
 \* 55203 55283: gap of unknown length  
 \* 55283 57891: contig of 2609 bp in length  
 \* 57892 57971: gap of unknown length  
 \* 57972 59715: contig of 1744 bp in length  
 \* 59716 59795: gap of unknown length  
 \* 59796 63092: contig of 3297 bp in length  
 \* 63093 63172: gap of unknown length  
 \* 63173 66602: contig of 3430 bp in length  
 \* 66603 66682: gap of unknown length  
 \* 66683 70903: contig of 4221 bp in length  
 \* 70904 70983: gap of unknown length  
 \* 70984 74368: contig of 3385 bp in length  
 \* 74369 74448: gap of unknown length  
 \* 74449 80276: contig of 5828 bp in length  
 \* 80277 80356: gap of unknown length  
 \* 80357 80708: contig of 348 bp in length  
 \* 80709 80784: gap of unknown length  
 \* 80785 81593: contig of 809 bp in length  
 \* 81594 81673: gap of unknown length  
 \* 81674 82415: contig of 741 bp in length  
 \* 82415 82495: gap of unknown length  
 \* 82495 82955: contig of 460 bp in length  
 \* 83035 83035: gap of unknown length  
 \* 83035 83456: contig of 422 bp in length  
 \* 83457 83536: gap of unknown length  
 \* 83537 84100: contig of 564 bp in length  
 \* 84101 84180: gap of unknown length  
 \* 84181 84685: contig of 505 bp in length  
 \* 84686 84765: gap of unknown length  
 \* 84766 85313: contig of 548 bp in length  
 \* 85314 85393: gap of unknown length  
 \* 85394 86028: contig of 635 bp in length  
 \* 86029 86108: gap of unknown length  
 \* 86109 86727: contig of 619 bp in length  
 \* 86728 86807: gap of unknown length  
 \* 86808 87354: contig of 547 bp in length  
 \* 87355 87434: gap of unknown length  
 \* 87435 88012: contig of 578 bp in length  
 \* 88013 88092: gap of unknown length  
 \* 88093 88532: contig of 440 bp in length  
 \* 88533 88612: gap of unknown length  
 \* 88613 89345: contig of 733 bp in length  
 \* 89346 89425: gap of unknown length  
 \* 89426 90007: contig of 582 bp in length  
 \* 90008 90087: gap of unknown length  
 \* 90088 90949: contig of 862 bp in length  
 \* 90950 91023: gap of unknown length  
 \* 91030 91496: contig of 467 bp in length  
 \* 91497 91576: gap of unknown length  
 \* 91577 92074: contig of 498 bp in length  
 \* 92075 92154: gap of unknown length  
 \* 92155 92748: contig of 594 bp in length  
 \* 92749 92828: gap of unknown length  
 \* 92830 93411: contig of 583 bp in length

\* 93412 93491: gap of unknown length  
 \* 93492 94063: contig of 572 bp in length  
 \* 94064 94143: gap of unknown length

Query Match 0.9%; Score 20; DB 2; Length 131813;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCAATGAATTAAGTA 22  
 |||||  
 Db 10083 GCAATGAATTAAGTA 10102

RESULT 10  
 AC120223/c

LOCUS AC120223 136241 bp DNA linear HTG 18-JUL-2002  
 DEFINITION Rattus norvegicus clone CH230-307M22, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 50 unordered pieces.  
 AC120223  
 AC120223.2 GI:21746475  
 VERSION  
 HTG: HTGS, PHASE1.  
 KEYWORDS Norway rat.  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 136241)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,  
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenwo,S., Ogih,M., Okwionu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,  
 Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tameris,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,O.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstock,G. and Gibbs,R.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 136241)  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-MAY-2002) Human Genome Sequencing Center, Department



of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 136241)  
 AUTHOR: Worley, K.C.  
 TITLE: Direct Submission  
 JOURNAL: Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 COMMENT: On Jul 14, 2002 this sequence version replaced gi:20452765.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 Center project name: GUML  
 Center clone name: CH230-307M22  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 92457 bases at least Q40  
 Consensus quality: 97846 bases at least Q30  
 Consensus quality: 101277 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 50 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 \*  
 1 1237: contig of 1237 bp in length  
 1238 1337: gap of unknown length  
 1338 2477: contig of 1140 bp in length  
 2478 2577: gap of unknown length  
 2578 3773: contig of 1196 bp in length  
 3774 3873: gap of unknown length  
 3874 5333: contig of 1460 bp in length  
 5334 5433: gap of unknown length  
 5434 6516: contig of 1083 bp in length  
 6517 6616: gap of unknown length  
 6617 7797: contig of 1181 bp in length  
 7798 7897: gap of unknown length  
 7898 9520: contig of 1623 bp in length  
 9521 9620: gap of unknown length  
 9621 11451: contig of 1831 bp in length  
 11452 11551: gap of unknown length  
 11552 12773: contig of 1222 bp in length  
 12774 12873: gap of unknown length  
 12874 14508: contig of 1633 bp in length  
 14509 14507: gap of unknown length  
 14508 15953: contig of 1347 bp in length  
 15954 16053: gap of unknown length  
 16054 18349: contig of 2296 bp in length  
 18350 18449: gap of unknown length  
 18450 19975: contig of 1527 bp in length  
 19977 20076: gap of unknown length  
 20077 21148: contig of 1072 bp in length  
 21149 21248: gap of unknown length  
 21249 22929: contig of 1681 bp in length  
 22930 23029: gap of unknown length  
 23030 25641: contig of 2612 bp in length  
 25642 25741: gap of unknown length  
 28054 28154: contig of 2313 bp in length  
 28155 30379: contig of 2225 bp in length  
 30380 30479: gap of unknown length  
 32748: contig of 2269 bp in length  
 32749 32848: gap of unknown length

32849 35106: contig of 2258 bp in length  
 35107 35206: gap of unknown length  
 35207 37048: contig of 1842 bp in length  
 37049 37148: gap of unknown length  
 37149 39266: contig of 2118 bp in length  
 39267 39366: gap of unknown length  
 39367 40899: contig of 1433 bp in length  
 40900 43019: contig of 2120 bp in length  
 43020 43119: gap of unknown length  
 43120 45158: contig of 2039 bp in length  
 45159 45258: gap of unknown length  
 45259 47295: contig of 2037 bp in length  
 47296 47395: gap of unknown length  
 47396 48834: contig of 1439 bp in length  
 48835 48934: gap of unknown length  
 48935 51435: contig of 2501 bp in length  
 51436 51535: gap of unknown length  
 51536 54235: contig of 2700 bp in length  
 54236 54335: gap of unknown length  
 54336 57329: contig of 2994 bp in length  
 57330 57429: gap of unknown length  
 57430 59496: contig of 2067 bp in length  
 59497 59596: gap of unknown length  
 59597 61948: contig of 2352 bp in length  
 61949 62048: gap of unknown length  
 62049 63956: contig of 1908 bp in length  
 63957 64056: gap of unknown length  
 64057 66236: contig of 2180 bp in length  
 66237 66336: gap of unknown length  
 66337 69688: contig of 3352 bp in length  
 69689 69788: gap of unknown length  
 69789 72181: contig of 2393 bp in length  
 72182 72281: gap of unknown length  
 72282 75455: contig of 3174 bp in length  
 75456 75555: gap of unknown length  
 75556 78895: contig of 3340 bp in length  
 78896 78995: gap of unknown length  
 78996 82341: contig of 3346 bp in length  
 82342 82441: gap of unknown length  
 82442 84795: contig of 2355 bp in length  
 84797 84896: gap of unknown length  
 84897 88837: contig of 3941 bp in length  
 88838 88937: gap of unknown length  
 88938 93012: contig of 4075 bp in length  
 93013 93112: gap of unknown length  
 93113 97714: contig of 4602 bp in length  
 97715 97814: gap of unknown length  
 97815 100110: contig of 2296 bp in length  
 100111 100210: gap of unknown length  
 100211 105379: contig of 5169 bp in length  
 105380 105479: gap of unknown length  
 105480 111300: contig of 5821 bp in length  
 111301 111400: gap of unknown length  
 111401 116430: contig of 5030 bp in length  
 116431 116530: gap of unknown length  
 116531 121578: contig of 5048 bp in length  
 121579 121678: gap of unknown length  
 121679 128057: contig of 6379 bp in length  
 128058 128157: gap of unknown length  
 128158 136241: contig of 8084 bp in length.  
 FEATURES  
 Location/Qualifiers  
 1. .136241  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-307M22"  
 BASE COUNT 39106 a 27003 c 27430 g 37671 t 5031 others  
 ORIGIN  
 Query Match 0.9%; Score 20; DB 2; Length 136241;  
 Best Local Similarity 100.0%; Pred. NO. 44;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

* 29886 31450: contig of 1565 bp in length
* 31451 31550: gap of unknown length
* 31551 33272: contig of 1722 bp in length
* 33273 33372: gap of unknown length
* 33373 34554: contig of 1182 bp in length
* 34555 34654: gap of unknown length
* 34655 35908: contig of 1234 bp in length
* 35909 37552: gap of unknown length
* 37553 37652: gap of unknown length
* 37653 39518: contig of 1866 bp in length
* 39519 39618: gap of unknown length
* 39619 41678: contig of 2060 bp in length
* 41679 41778: gap of unknown length
* 41779 43376: contig of 1598 bp in length
* 43377 43476: gap of unknown length
* 43477 45110: contig of 1634 bp in length
* 45111 45210: gap of unknown length
* 45211 46627: contig of 1417 bp in length
* 46628 46727: gap of unknown length
* 46728 48564: contig of 1837 bp in length
* 48565 48664: gap of unknown length
* 48665 49903: contig of 1239 bp in length
* 49904 50003: gap of unknown length
* 50004 51077: contig of 1074 bp in length
* 51078 51177: gap of unknown length
* 51178 52965: contig of 1788 bp in length
* 52966 53065: gap of unknown length
* 53066 55285: contig of 2220 bp in length
* 55286 55385: gap of unknown length
* 55386 56530: contig of 1145 bp in length
* 56531 56630: gap of unknown length
* 56631 57808: contig of 1178 bp in length
* 57809 59912: contig of 2004 bp in length
* 59913 60012: gap of unknown length
* 60013 61628: contig of 1616 bp in length
* 61629 61728: gap of unknown length
* 61729 63702: contig of 1974 bp in length
* 63703 63802: gap of unknown length
* 63803 65518: contig of 1716 bp in length
* 65519 65618: gap of unknown length
* 65619 67487: contig of 1869 bp in length
* 67488 67587: gap of unknown length
* 67588 69222: contig of 1635 bp in length
* 69223 69322: gap of unknown length
* 69323 71614: contig of 2292 bp in length
* 71615 74337: contig of 2623 bp in length
* 74338 74437: gap of unknown length
* 74438 76038: contig of 1601 bp in length
* 76039 76138: gap of unknown length
* 76139 78343: contig of 2205 bp in length
* 78344 78444: gap of unknown length
* 78444 80973: contig of 2530 bp in length
* 80974 81073: gap of unknown length
* 81074 83111: contig of 2038 bp in length
* 83112 83211: gap of unknown length
* 83212 85690: contig of 2479 bp in length
* 85691 85790: gap of unknown length
* 85791 88004: contig of 2214 bp in length

```

```

Query Match      0.9%; Score 20; DB 2; Length 150505;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 504 TGCAGACGAGAACCACTCT 523
|||||
Db 31654 TGCAGACGAGAACCACTCT 31635

```

```

RESULT 12
AL391294/c

```

```

LOCUS      AL391294      152120 bp      DNA      linear      PRI 06-DEC-2001
DEFINITION Human DNA sequence from clone Rpl1-196G4 on chromosome X Contains
              part of a novel gene and a CpG island, complete sequence.
ACCESSION  AL391294
VERSION    AL391294.16  GI:11121074
KEYWORDS   HTG; CpG island.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 152120)
AUTHORS   Clark,S.
TITLE     Direct Submission
JOURNAL   Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Nov 8, 2000 this sequence version replaced gi:11064146.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            This sequence is the entire insert of clone Rpl1-196G4 The true
            right end of clone Rpl3-104H2 is at 71449 in this sequence. This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. Rpl1-196G4 is from
            the library RPCI-11.1 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACE3.6.
FEATURES             Location/Qualifiers
     source            1..152120
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="X"
                        /clone="Rpl1-196G4"
                        /clone_lib="RPCI-11.1"
     repeat_region     1..1563
                        /note="TIGER1 repeat: matches 1..1558 of consensus"
     repeat_region     1666..1815
                        /note="75 copies 2 mer tt 58% conserved"
     repeat_region     2485..2668
                        /note="MER5A repeat: matches 1..184 of consensus"
     repeat_region     3055..3100
                        /note="23 copies 2 mer tt 87% conserved"
     repeat_region     3809..4297
                        /note="MER9 repeat: matches 1..511 of consensus"
     repeat_region     5346..5399
                        /note="27 copies 2 mer ac 98% conserved"
     repeat_region     5507..6137
                        /note="L1M1 repeat: matches -1389..-744 of consensus"
     repeat_region     6110..6247
                        /note="L1M1 repeat: matches -1186..-1037 of consensus"
     repeat_region     6231..8974
                        /note="L1M1 repeat: matches -523..-2714 of consensus"
     misc_feature      complement(8868..9026)
                        /note="match: GSS: Em:AZ392746"
     misc_feature      8952..9312

```

```
/note="match: GSS: Em:AQ236903"
complement(8971..9176)
/note="match: GSS: Em:AQ52982 Em:AQ568449"
8971..9159
/note="match: SFS: Em:Af002096"
complement(9021..9179)
/note="match: SFS: Em:G09176"
9028..9179
/note="AluJb repeat: matches 138..289 of consensus"
9180..12619
/note="L1M2 repeat: matches 2325..5786 of consensus"
12614..14953
/note="L1P13 repeat: matches 3768..6152 of consensus"
14961..15487
/note="L1M2 repeat: matches 5769..6303 of consensus"
complement(15532..17999)
/gene="bA19664.1"
complement(<15532..17999)
/gene="bA19664.1"
/product="bA19664.1 (novel protein)"
/note="match: cDNAs: Em:AK024111 Em:AK001535 Em:AL050030
match: ESTs: Em:AI252657 Em:BE673395 Em:AA947819
Em:BF142469 Em:AW883384 Em:BE005185 Em:BE180722
Em:BF141331 Em:AW892817 Em:BE326061 Em:BF139073 Em:F12610
Em:BE707336 Em:AA968512 Em:AA634946 Em:F08032 Em:BE781553
Em:BE018659 Em:W56397 Em:BE174649 Em:AI075935 Em:BE992345
Em:BF144949 Em:BE916748 Em:H17605 Em:AI694583 Em:AW524670
Em:AI798375 Em:BE005114 Em:AI110805 Em:BF140831
Em:AW369780 Em:AW885772 Em:AW885736 Em:BF056380
Em:AI065144 Em:RI9847 Em:AW889619 Em:BE613917 Em:AW900587
Em:BE178770 Em:BE171589 Em:BE881716 Em:AA243338
Em:BE770557 Em:AA603412 Em:AA280341 Em:AW293265
Em:AW181892 Em:AW885540 Em:AW517792 Em:AA262381 Em:T06923
Em:AW993831 Em:BE163571 Em:BE170628 Em:BE709135
Em:BE564899 Em:BE175078 Em:F12118 Em:BE465085 Em:AI250884
Em:AA837637 Em:AA010153 Em:AA987568 Em:AW771325
Em:AW780434 Em:AW62521 Em:T25165 Em:AA242913 Em:W31490
Em:AW892807 Em:AF075304 Em:BE270146 Em:R57609 Em:AW879991
Em:BE915168 Em:AA598509 Em:BE827473 Em:AI572844
Em:BE827475 Em:AI370669 Em:AW900704 Em:AW270335
Em:BE569882 Em:AW71253 Em:AW660653"
/evidence="not_experimental"
15801..16081
/note="AlusC repeat: matches 1..281 of consensus"
complement(16083..16344)
/gene="bA19664.1"
/note="match: GSS: Em:A2271635"
complement(17589..17996)
/gene="bA19664.1"
/note="match: GSS: Em:AQ730707"
complement(18204..18563)
/note="match: GSS: Em:AQ730707"
19122..19525
/note="CpG island"
/evidence="not_experimental"
21236..21293
/note="MIR repeat: matches 85..142 of consensus"
21550..21678
/note="MLN1A2 repeat: matches 220..349 of consensus"
21881..24760
/note="L1P2 repeat: matches 3013..6152 of consensus"
21804..21855
/note="Sequence from overlapping clone AL359775. Assembly
confirmed by restriction digest"
24754..25642
/note="L1P repeat: matches 1560..2455 of consensus"
25634..26219
/note="L1P3 repeat: matches 8..416 of consensus"
26225..26480
/note="L1MD1 repeat: matches 5566..5825 of consensus"
26518..27573
/note="L1MEC repeat: matches 358..1420 of consensus"
27597..27650

/note="THE1C-internal repeat: matches 1524..1576 of
consensus"
27655..28025
/note="THE1C repeat: matches 1..371 of consensus"
28026..28120
/note="FLAM_A repeat: matches 29..123 of consensus"
28186..28209
/note="12 copies 2 mer ac 100% conserved"
28331..28775
/note="MLT2A repeat: matches 9..453 of consensus"
29122..29426
/note="Alusx repeat: matches 5..305 of consensus"
29427..30036
/note="L1P8 repeat: matches 10..602 of consensus"
30037..30502
/note="L1P8 repeat: matches 5696..6161 of consensus"
30572..31103
/note="L1MD2 repeat: matches 5575..6116 of consensus"
31104..31624
/note="L1M4 repeat: matches -2..556 of consensus"
31629..34385
/note="L1P repeat: matches 1818..4658 of consensus"
34390..34461
/note="24 copies 3 mer tat 68% conserved"
34392..34463
/note="12 copies 6 mer ttatta 68% conserved"
34500..34936
/note="L1P8 repeat: matches 5724..6163 of consensus"
35012..37203
/note="L1P12 repeat: matches -355..1833 of consensus"
37383..37682
/note="L1M2 repeat: matches -614..-321 of consensus"
37755..38071
/note="L1M4 repeat: matches 4330..4650 of consensus"
38052..38277
/note="L1P12 repeat: matches 1323..1554 of consensus"
38407..39569
/note="L1P repeat: matches 2047..3228 of consensus"
39571..41509
/note="L1P15 repeat: matches 4193..6148 of consensus"
41502..42355
/note="L1 repeat: matches 3785..4644 of consensus"
42366..48504
/note="L1P4 repeat: matches 1..6141 of consensus"
49131..49187
/note="MER93 repeat: matches 17..82 of consensus"
49302..49911
/note="L1P5 repeat: matches 5529..6138 of consensus"
49921..50335
/note="L1P5 repeat: matches 5135..5552 of consensus"
50392..51778
/note="L1P4 repeat: matches 2176..3577 of consensus"
51779..52334
/note="L1R1 repeat: matches 194..785 of consensus"

Query Match 0.9%; Score 20; DB 9; Length 152120;
Best Local Similarity 100.0%; Pred.No.45;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 485 AACTGGCCAAAAGAAAAT 504
|||||
Db 139650 AACTGGCCAAAAGAAAAT 139631

RESULT 13
AC019024/c AC019024 linear HTG 14-MAR-2002
LOCUS Homo sapiens chromosome 11 clone RP11-478E10 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 3 ordered pieces.
ACCESSION AC019024
VERSION AC019024.5 GI:19424608
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens.
```

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.

## REFERENCE

1 (bases 1 to 160279)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

## JOURNAL

Unpublished

## AUTHORS

2 (bases 1 to 160279)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bida, F., Boguslavskiy, L., Bouckghalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArlanno, K., Dewar, K., Domino, M., Doyle, M., Fenesfor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, B., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (29-DEC-1999)

Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

3 (bases 1 to 160279)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckghalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chokaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArlanno, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (14-MAR-2002)

Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Mar 14, 2002 this sequence version replaced gi:19387766.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L5298

Center clone name: 478\_E\_10

----- Summary Statistics

Sequencing vector: M13; M77815; 26% of reads

Sequencing vector: Plasmid; n/a; 74% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 159553 bases at least Q40

Consensus quality: 159829 bases at least Q30

Consensus quality: 159947 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 160079; sum-of-ctgigs

Quality coverage: 19.1 in Q20 bases; agarose-fp

Quality coverage: 19.5 in Q20.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 77856: contig of 77856 bp in length

\* 77857 77956: gap of 100 bp

\* 77957 79239: contig of 1283 bp in length

\* 79240 79339: gap of 100 bp

\* 79340 160279: contig of 80940 bp in length.

\* Location/Qualifiers

1..160279

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosomes="11"

/map="11"

/clone="RP11-478E10"

/clone\_lib="RP11-11 Human Male BAC"

1..77856

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left

77957..79239

/note="assembly\_fragment"

79340..160279

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right

BASE COUNT 42570 a 40290 c 38390 g 38828 t 201 others

ORIGIN

Query Match 0.9% Score 20; DB 2; Length 160279;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1135 CTCGAAAGGCTTATAGGA 1154

|||||

Db 66837 CTCGAAAGGCTTATAGGA 66818

|||||

RESULT 14

AC009253

LOCUS

DEFINITION

Drosophila melanogaster, chromosome 2L, region 39A-39A, BAC clone

BACR24F17, complete sequence.

AC009253

VERSION

AC009253.16 GI:16798933

HTG.

SOURCE

ORGANISM

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 160282)

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,

Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,

Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busan, D.A.,

Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M.,

Donson, K., Dorsett, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D.,

Ferrera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,

Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,

Ihegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A.,

McIntosh, T.C., Moy, M., Murphy, B., Nelson, K.A., Nunoo, J.,

Pacleb, J., Paragas, V., Park, S., Pfaffner, B.,

Patel, S., Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Patel, S., Pfaffner, B.,

Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 2L, region 39A-39A

Unpublished

2 (bases 1 to 160282)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomoton, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacieb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zietan, L.L. and Rubin, G.M.

Direct Submission

Submitted (09-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Nov 8, 2001 this sequence version replaced gi:7264755.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgpf@fruitfly.berkeley.edu](mailto:bdgpf@fruitfly.berkeley.edu).

Location/Qualifiers

1. 160282

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="2L"

/map="39A-39A"

/clone="BACR24F17 (D1038)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

pBACE3.6)"

BASE COUNT 47892 a 33440 c 31695 g 47255 t

ORIGIN

Query Match 0.98; Score 20; DB 3; Length 160282;

Best Local Similarity 100.0%; Pred. NO. 45;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCAATGAAATTAAGTA 22

|||||

Db 3371 GCAATGAAATTAAGTA 3390

RESULT 15

AC098197

LOCUS AC098197 163706 bp DNA linear HTG 12-JUL-2002

DEFINITION Rattus norvegicus clone CH230-9714, \*\*\* SEQUENCING IN PROGRESS \*\*\*,

65 unordered pieces.

AC098197

AC098197 5 GI:21729815

VERSION HTG; HTGS-PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 163706)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amarutunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brivela, M., Brown, E., Brown, M., Bryant, N.P., Buhay, J., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, T., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamarisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 163706)

Worley, K.C.

Direct Submission

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 163706)

Worley, K.C.

Direct Submission

Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jul 11, 2002 this sequence version replaced gi:17973759.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GIKF

Center clone name: CH230-9714

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 116803 bases at least Q40

Consensus quality: 121625 bases at least Q30

Consensus quality: 125275 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 65 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will

\* be preserved.  
1  
1248: contig of 1248 bp in length  
1348: gap of unknown length  
1349: contig of 1210 bp in length  
2558: gap of unknown length  
2559: gap of unknown length  
2659: contig of 1260 bp in length  
3919: gap of unknown length  
4019: contig of 1543 bp in length  
5561: gap of unknown length  
5562: contig of 1446 bp in length  
5662: gap of unknown length  
7107: contig of 1446 bp in length  
7207: gap of unknown length  
8207: contig of 1000 bp in length  
8307: gap of unknown length  
8308: contig of 1426 bp in length  
9733: gap of unknown length  
9734: gap of unknown length  
9834: contig of 1021 bp in length  
10854: gap of unknown length  
10954: gap of unknown length  
12152: contig of 1198 bp in length  
12153: gap of unknown length  
12253: contig of 1551 bp in length  
13803: gap of unknown length  
13804: contig of 1080 bp in length  
14983: gap of unknown length  
15083: gap of unknown length  
16304: contig of 1221 bp in length  
16404: gap of unknown length  
17407: contig of 1003 bp in length  
17507: gap of unknown length  
18723: contig of 1216 bp in length  
18724: gap of unknown length  
18823: contig of 1050 bp in length  
18824: gap of unknown length  
19773: gap of unknown length  
19774: contig of 1377 bp in length  
21350: gap of unknown length  
21450: gap of unknown length  
22583: contig of 1133 bp in length  
22584: gap of unknown length  
22684: contig of 1168 bp in length  
23951: gap of unknown length  
23952: contig of 2027 bp in length  
25978: gap of unknown length  
26078: contig of 1619 bp in length  
27697: gap of unknown length  
27698: contig of 1086 bp in length  
28883: gap of unknown length  
28983: contig of 1034 bp in length  
30017: gap of unknown length  
30117: gap of unknown length  
31386: contig of 1269 bp in length  
31486: gap of unknown length  
33565: contig of 2079 bp in length  
33665: gap of unknown length  
35234: contig of 1569 bp in length  
35334: gap of unknown length  
35235: contig of 1623 bp in length  
36957: gap of unknown length  
37057: contig of 1370 bp in length  
38427: gap of unknown length  
38527: gap of unknown length  
38528: contig of 1380 bp in length  
40007: gap of unknown length  
41539: contig of 1532 bp in length  
41639: gap of unknown length  
43413: contig of 1774 bp in length  
43513: gap of unknown length  
45389: contig of 1876 bp in length  
45489: gap of unknown length  
47738: contig of 2249 bp in length  
47838: gap of unknown length  
48892: contig of 2054 bp in length  
49992: gap of unknown length  
49993: contig of 2485 bp in length  
52577: gap of unknown length  
52578: contig of 2232 bp in length  
54809: gap of unknown length  
54910: contig of 1242 bp in length  
56151: gap of unknown length  
56251: gap of unknown length

\* 56252 58273: contig of 2022 bp in length  
\* 58274 58373: gap of unknown length  
\* 58374 61788: contig of 3415 bp in length  
\* 61789 61888: gap of unknown length  
\* 61889 64074: contig of 2186 bp in length  
\* 64075 64174: gap of unknown length  
\* 64175 67262: contig of 3088 bp in length  
\* 67263 67362: gap of unknown length  
\* 67363 70423: contig of 3061 bp in length  
\* 70424 70523: gap of unknown length  
\* 70524 74022: contig of 3499 bp in length  
\* 74023 74122: gap of unknown length  
\* 74123 76188: contig of 2066 bp in length  
\* 76189 76288: gap of unknown length  
\* 76289 78899: contig of 2611 bp in length  
\* 78900 78999: gap of unknown length  
\* 79000 80837: contig of 1838 bp in length  
\* 80838 80937: gap of unknown length  
\* 80938 83128: contig of 2191 bp in length  
\* 83129 83228: gap of unknown length  
\* 83229 85835: contig of 2607 bp in length  
\* 85836 85935: gap of unknown length  
\* 85936 88399: contig of 2464 bp in length  
\* 88400 88499: gap of unknown length  
\* 88500 92442: contig of 3943 bp in length  
\* 92443 92542: gap of unknown length  
\* 92543 96214: contig of 3672 bp in length  
\* 96215 96314: gap of unknown length  
\* 96315 99043: contig of 2729 bp in length  
\* 99044 99144: gap of unknown length  
\* 99145 103453: contig of 4310 bp in length  
\* 103454 103553: gap of unknown length  
\* 103554 106323: contig of 2770 bp in length  
\* 106324 106423: gap of unknown length  
\* 106424 110557: contig of 4134 bp in length

Query Match 0.9%; Score 20; DB 2; Length 163706;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1160 CCAAGGCCAACCGTGAGATG 1179  
Db 136347 CCAAGGCCAACCGTGAGATG 136366  
|||||

Search completed: May 23, 2003, 08:42:14  
Job time : 5892 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 07:01:32 ; Search time 193 Seconds  
(without alignments)  
14634.553 Million cell updates/sec

Title: US-10-008-355-1  
Perfect score: 2139  
Sequence: 1 atgcaaatgaaattaaaag.....aagagctgaagttgatctaa 2139

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 828747 seqs, 560231138 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Published Applications\_NA.\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMP.seq.\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMP.seq.\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMP.seq.\*
  - 7: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMP.seq.\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMP.seq.\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMP.seq.\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMP.seq.\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2139	100.0	2139	9	US-10-008-355-1
2	19	0.9	490	9	US-09-918-995-37887
3	19	0.9	505	9	US-09-918-995-37842
4	19	0.9	4163	10	US-09-954-456-528
5	18	0.8	203	10	US-09-960-352-10103
6	18	0.8	252	10	US-09-974-300-7766
7	18	0.8	355	10	US-09-960-352-12702
8	18	0.8	408	10	US-09-960-352-9104
9	18	0.8	410	10	US-09-960-352-10777
10	18	0.8	417	10	US-09-960-352-5137
11	18	0.8	418	10	US-09-960-352-11256
12	18	0.8	427	10	US-09-960-352-12448
13	18	0.8	466	10	US-09-960-352-16022
14	18	0.8	499	9	US-09-918-995-1375
15	18	0.8	2699	8	US-08-834-666A-3
16	18	0.8	2798	8	US-08-834-666A-1
17	18	0.8	2915	8	US-08-834-666A-5
18	18	0.8	3995	10	US-09-919-497-9
19	18	0.8	3995	10	US-09-919-497-9

c	20	18	0.8	7557	12	US-10-044-090-253	Sequence 253, App
c	21	17	0.8	264	9	US-10-046-935-657	Sequence 657, App
c	22	17	0.8	264	9	US-09-878-178-657	Sequence 657, App
c	23	17	0.8	264	9	US-10-146-502-657	Sequence 657, App
c	24	17	0.8	285	10	US-09-923-876-4571	Sequence 4571, App
c	25	17	0.8	287	10	US-09-960-352-478	Sequence 478, App
c	26	17	0.8	348	9	US-09-918-995-18569	Sequence 18569, A
c	27	17	0.8	373	10	US-09-880-107-614	Sequence 614, App
c	28	17	0.8	399	10	US-09-960-253-126	Sequence 126, App
c	29	17	0.8	402	9	US-09-918-995-3910	Sequence 3910, App
c	30	17	0.8	413	10	US-09-960-352-2414	Sequence 2414, App
c	31	17	0.8	470	9	US-09-918-995-25073	Sequence 25073, A
c	32	17	0.8	485	9	US-09-918-995-23928	Sequence 23928, A
c	33	17	0.8	495	10	US-09-783-590-8450	Sequence 8450, App
c	34	17	0.8	516	10	US-09-919-580-852	Sequence 852, App
c	35	17	0.8	600	10	US-09-917-800A-857	Sequence 857, App
c	36	17	0.8	606	9	US-09-774-639-50	Sequence 50, Appl
c	37	17	0.8	606	9	US-09-969-730-98	Sequence 98, Appl
c	38	17	0.8	623	9	US-09-883-152-29	Sequence 29, Appl
c	39	17	0.8	651	9	US-09-883-152-26	Sequence 26, Appl
c	40	17	0.8	687	9	US-09-774-639-106	Sequence 106, App
c	41	17	0.8	687	9	US-09-969-730-107	Sequence 107, App
c	42	17	0.8	861	10	US-09-452-598-166	Sequence 166, App
c	43	17	0.8	918	10	US-09-974-300-4593	Sequence 4593, App
c	44	17	0.8	981	9	US-09-887-527-12	Sequence 12, Appl
c	45	17	0.8	998	9	US-10-001-857-87	Sequence 87, Appl

ALIGNMENTS

RESULT 1  
US-10-008-355-1  
; Sequence 1, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 1  
; LENGTH: 2139  
; TYPE: DNA  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-1

Query Match	100.0%	Score	2139;	DB	9;	Length	2139;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	2139;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	ATGCAAAATGAAATTAATAAGTATTCTTCGAGACAGCCCTGCTGTGGTGGTCTTACAGG	60				
Db	1	ATGCAAAATGAAATTAATAAGTATTCTTCGAGACAGCCCTGCTGTGGTGGTCTTACAGG	60				
Qy	61	GTAGCCAAAGCCGAAAGGATGTGGCTCTCTCAACGAATCAATCAGGAGAATCTGGAT	120				
Db	61	GTAGCCAAAGCCGAAAGGATGTGGCTCTCTCAACGAATCAATCAGGAGAATCTGGAT	120				
Qy	121	CGAATGCTGAGCTCGGCTTTTACGCTCCCGTTGGATTGCGTCTACAGTTTCGACAAGCCG	180				
Db	121	CGAATGCTGAGCTCGGCTTTTACGCTCCCGTTGGATTGCGTCTACAGTTTCGACAAGCCG	180				
Qy	181	TCCATTGCCAATGCCGTGGTGTATCTTCGGTGGCGGATGATCCGGTATCACAGTGTCCGAT	240				
Db	181	TCCATTGCCAATGCCGTGGTGTATCTTCGGTGGCGGATGATCCGGTATCACAGTGTCCGAT	240				

QY 241 CAGGCGCTGATCTTTACCAACCACTGCGGATACGGTCTATTCAGAGCCAAAGCACG 300  
Db 241 CAGGCGCTGATCTTTACCAACCACTGCGGATACGGTCTATTCAGAGCCAAAGCACG 300  
QY 301 GTGGATCACGACTATCTGCGGGATGGTTTCTGTTCTCGCAGATGGGTGAGGAGCTTCG 360  
Db 301 GTGGATCACGACTATCTGCGCGATGGTTTCTGTTCTCGCAGATGGGTGAGGAGCTTCG 360  
QY 361 ATTCCGGGCTCTTCCCGTGAAGTATCTCGCAAGATCTGGAAGTAAACGGACAAAGTGAA 420  
Db 361 ATTCCGGGCTCTTCCCGTGAAGTATCTCGCAAGATCTGGAAGTAAACGGACAAAGTGAA 420  
QY 421 GGACAGCTCAAGGATATCTAGTACGAGATGAGCGTCTGCGCAAGCTCAGGAGTATGC 480  
Db 421 GGACAGCTCAAGGATATCTAGTACGAGATGAGCGTCTGCGCAAGCTCAGGAGTATGC 480  
QY 481 CAAGAACTGGCCAAAAGAAATGCAAGACGAGACCACTCTGCATCGTAGAGCCTTTC 540  
Db 481 CAAGAACTGGCCAAAAGAAATGCAAGACGAGACCACTCTGCATCGTAGAGCCTTTC 540  
QY 541 TATTCCAAACAAAGTATCTTCTCATCTCTACGATGTATTCAAGGACGTTTCGTATGSTA 600  
Db 541 TATTCCAAACAAAGTATCTTCTCATCTCTACGATGTATTCAAGGACGTTTCGTATGSTA 600  
QY 601 TTTGCTCCTCCAGCTCTGTAGTAACTTCGGAGCGGATACGGACAACTGGATGTGCGCG 660  
Db 601 TTTGCTCCTCCAGCTCTGTAGTAACTTCGGAGCGGATACGGACAACTGGATGTGCGCG 660  
QY 661 CGTCACAGCGGCGACTTCAGCGTATTCGGGTGTATGCCGGTCCGACAAACCGCGCGCC 720  
Db 661 CGTCACAGCGGCGACTTCAGCGTATTCGGGTGTATGCCGGTCCGACAAACCGCGCGCC 720  
QY 721 GAATACAGCAAGGACAAATAAACCCCTATAAGCCGCTTTACTTCGCTGCCGTATCCATGCA 780  
Db 721 GAATACAGCAAGGACAAATAAACCCCTATAAGCCGCTTTACTTCGCTGCCGTATCCATGCA 780  
QY 781 GGCTACAGGCTGACGACTATGCCATGACCATCGGTTTCCGGGACGATCGGATCGGTAC 840  
Db 781 GGCTACAGGCTGACGACTATGCCATGACCATCGGTTTCCGGGACGATCGGATCGGTAC 840  
QY 841 CTCACCTTTGGGGTGGAGATCGTATCGAAGGATCGTATCGAAGGATCGTATCGAAGT 900  
Db 841 CTCACCTTTGGGGTGGAGATCGTATCGAAGGATCGTATCGAAGGATCGTATCGAAGT 900  
QY 901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAAGCCATGAGCGAGATCAGGCTACCCGATC 960  
Db 901 CGCGGTATCAAGCAAGGCATCTGGAAGGAAAGCCATGAGCGAGATCAGGCTACCCGATC 960  
QY 961 AAATATCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAAATTCGATCGGTATGAC 1020  
Db 961 AAATATCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAAATTCGATCGGTATGAC 1020  
QY 1021 CGGGTCTCGCTCGTTCGATAGTATGCTGTAAGCGTCCGAGGAAAGAGCATTCGCA 1080  
Db 1021 CGGGTCTCGCTCGTTCGATAGTATGCTGTAAGCGTCCGAGGAAAGAGCATTCGCA 1080  
QY 1081 GACTGGATCGTAAAGCAAGGATGCTGCTATGGCGATGATTTGCTTCTCTCGAA 1140  
Db 1081 GACTGGATCGTAAAGCAAGGATGCTGCTATGGCGATGATTTGCTTCTCTCGAA 1140  
QY 1141 AAGCCTTAAAGCAAGGATGCTGCTATGGCGATGATTTGCTTCTCTCGAA 1200  
Db 1141 AAGCCTTAAAGCAAGGATGCTGCTATGGCGATGATTTGCTTCTCTCGAA 1200  
QY 1201 TTGGTGGTACCGAGGTTGCTTTTTCACAGTTTCCCAACGATTCGCTACAAATCCT 1260  
Db 1201 TTGGTGGTACCGAGGTTGCTTTTTCACAGTTTCCCAACGATTCGCTACAAATCCT 1260  
QY 1261 GATGCTCATCGGATTCCTCAATCGCTTGGAGCAAGTACAAAGACTACCTCCCGCTCG 1320  
Db 1261 GATGCTCATCGGATTCCTCAATCGCTTGGAGCAAGTACAAAGACTACCTCCCGCTCG 1320  
QY 1321 CTGACCGTAAGGTGCTGCCCGCATCTCGATATTGTACGCCGGCGTATTCCTCGCCGAC 1380

Db 1321 CTGACCGTAAGGTGCTGCCCGCATCTCGATATTGTACGCCGGCGTATTCCTCGCCGAC 1380  
QY 1381 AAGCTCCCCGATATATTCAAGATGTAATCGACAAGAAATTCAAAGCGCACAGCAAG 1440  
Db 1381 AAGCTCCCCGATATATTCAAGATGTAATCGACAAGAAATTCAAAGCGCACAGCAAG 1440  
QY 1441 TATGCAAGCTTCGATATTGCAAGAGTGTGGTTCCTTTATAGCGACAAAGTTCCATGCCATG 1500  
Db 1441 TATGCAAGCTTCGATATTGCAAGAGTGTGGTTCCTTTATAGCGACAAAGTTCCATGCCATG 1500  
QY 1501 CTAAAGTCCATGACAAAGGAAAGTTTGCACAGGCTATTCGAGAAAGATCCGGCAGTAGAG 1560  
Db 1501 CTAAAGTCCATGACAAAGGAAAGTTTGCACAGGCTATTCGAGAAAGATCCGGCAGTAGAG 1560  
QY 1561 CTTTCCAAAGAGCGTATAGTCTGCTCGCGCTATTTCAGGCCGATCGATGGCCAAATGCC 1620  
Db 1561 CTTTCCAAAGAGCGTATAGTCTGCTCGCGCTATTTCAGGCCGATCGATGGCCAAATGCC 1620  
QY 1621 TATGCCATTCAGAAAGGCGGCTTTTCTTTCCCGGTTTGGCTGAGATGTACCCCGGA 1680  
Db 1621 TATGCCATTCAGAAAGGCGGCTTTTCTTTCCCGGTTTGGCTGAGATGTACCCCGGA 1680  
QY 1681 CGTCTCTCGCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740  
Db 1681 CGTCTCTCGCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGGCTCCATCAAGGGA 1740  
QY 1741 TATGAACCGGAGGCGGCTGCTGTACAACATATCATACGACAGCAAGGCGGTATTGAG 1800  
Db 1741 TATGAACCGGAGGCGGCTGCTGTACAACATATCATACGACAGCAAGGCGGTATTGAG 1800  
QY 1801 AAGCAGGATCCTTAAGAGCGATGAGTTTGGCGTACAGAGAAATATCCTCGACCTTCCCGC 1860  
Db 1801 AAGCAGGATCCTTAAGAGCGATGAGTTTGGCGTACAGAGAAATATCCTCGACCTTCCCGC 1860  
QY 1861 ACCAAACATATGCTGCTATGCGGAGAACGGTACGATCCTATCGCTTTCCTATCGAAC 1920  
Db 1861 ACCAAACATATGCTGCTATGCGGAGAACGGTACGATCCTATCGCTTTCCTATCGAAC 1920  
QY 1921 AAGCAGATCAGGCGGTAACCTCCGCTAGCGGCTATTCGATAGAACGCGCTGCTATC 1980  
Db 1921 AAGCAGATCAGGCGGTAACCTCCGCTAGCGGCTATTCGATAGAACGCGCTGCTATC 1980  
QY 1981 GGTCTTCTTCGATGCGCAACTGGGAAGCTATGAGTGTGACATCGAGTTCGAACCCGAT 2040  
Db 1981 GGTCTTCTTCGATGCGCAACTGGGAAGCTATGAGTGTGACATCGAGTTCGAACCCGAT 2040  
QY 2041 CTGAGCGGACATCAGCGTGGACATCCGCTACGTTCTTCTTCATGATTGACAAATGGGT 2100  
Db 2041 CTGAGCGGACATCAGCGTGGACATCCGCTACGTTCTTCTTCATGATTGACAAATGGGT 2100  
QY 2101 CAGTGGCCCCGCTCTCATCCAAAGAGCTGAAGTTGATCTAA 2139  
Db 2101 CAGTGGCCCCGCTCTCATCCAAAGAGCTGAAGTTGATCTAA 2139

RESULT 2  
US-09-918-995-37887  
; Sequence 37887, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 37887  
; LENGTH: 490

```

; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 528
; LENGTH: 4163
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-528

Query Match          0.9%; Score 19; DB 10; Length 4163;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 CAGGGCTGATCTTTACCA 259
      |||||
DB 3135 CAGGGCTGATCTTTACCA 3153

RESULT 5
US-09-960-352-10103
; Sequence 10103; Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10103
; LENGTH: 203
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 43-LIB34-066-Q1-El-C12
US-09-960-352-10103

Query Match          0.8%; Score 18; DB 10; Length 203;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1160 CCAGGCCAACCGTGAGA 1177
      |||||
DB 57 CCAAGGCCAACCGTGAGA 74

RESULT 6
US-09-974-300-7766/c
; Sequence 7766; Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: Expression
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27

```

; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7766  
; LENGTH: 252  
; TYPE: DNA  
; ORGANISM: Bacillus clausii  
US-09-974-300-7766

Query Match 0.8%; Score 18; DB 10; Length 252;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 639 TACGGACAACCTGGATGTG 656  
|||||  
Db 159 TACGGACAACCTGGATGTG 142

## RESULT 7

US-09-960-352-12702  
; Sequence 12702, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 12702  
; LENGTH: 355  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 54-LIB34-010-Q1-E1-F10  
US-09-960-352-12702

Query Match 0.8%; Score 18; DB 10; Length 355;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1160 CCAAGGCCAACCGTGAGA 1177  
|||||  
Db 335 CCAAGGCCAACCGTGAGA 352

## RESULT 8

US-09-960-352-9104  
; Sequence 9104, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 9104  
; LENGTH: 408  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 39-LIB3058-039-Q1-K1-B4  
US-09-960-352-9104

Query Match 0.8%; Score 18; DB 10; Length 408;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1160 CCAAGGCCAACCGTGAGA 1177  
|||||  
Db 384 CCAAGGCCAACCGTGAGA 401

## RESULT 9

US-09-960-352-10777  
; Sequence 10777, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 10777  
; LENGTH: 410  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (383)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: 46-LIB34-016-Q1-E1-D6  
US-09-960-352-10777

Query Match 0.8%; Score 18; DB 10; Length 410;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1160 CCAAGGCCAACCGTGAGA 1177  
|||||  
Db 386 CCAAGGCCAACCGTGAGA 403

## RESULT 10

US-09-960-352-5137  
; Sequence 5137, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 5137  
; LENGTH: 417  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 22-LIB34-030-Q1-E1-F5  
US-09-960-352-5137

Query Match 0.8%; Score 18; DB 10; Length 417;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1160 CCAAGGCCAACCGTGAGA 1177  
|||||  
Db 117 CCAAGGCCAACCGTGAGA 134

## RESULT 11

US-09-960-352-11256  
; Sequence 11256, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 11256  
; LENGTH: 418  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 48-LIB34-016-Q1-E1-D8  
US-09-960-352-11256

Query Match 0.8%; Score 18; DB 10; Length 418;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1160 CCAAGGCCAACCGTGAGA 1177  
|||||  
Db 346 CCAAGGCCAACCGTGAGA 363  
|||||

RESULT 12  
US-09-960-352-12448  
; Sequence 12448, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960.352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 12448  
; LENGTH: 427  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 53-LIB34-010-Q1-E2-F10  
US-09-960-352-12448

Query Match 0.8%; Score 18; DB 10; Length 427;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1160 CCAAGGCCAACCGTGAGA 1177  
|||||  
Db 335 CCAAGGCCAACCGTGAGA 352  
|||||

RESULT 13  
US-09-864-761-6022/c  
; Sequence 6022, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864.761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24363.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/006666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 6022  
; LENGTH: 466  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC004159.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9  
; OTHER INFORMATION: EXPRESSED IN HEAT, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6  
US-09-864-761-6022

Query Match 0.8%; Score 18; DB 10; Length 466;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 486 ACTGGCCAAAAAGAAAA 503  
|||||  
Db 135 ACTGGCCAAAAAGAAAA 118  
|||||

RESULT 14  
US-09-960-352-14538  
; Sequence 14538, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 14538  
LENGTH: 479  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 62-LIB3058-022-Q1-K1-H6  
US-09-960-352-14538

Query Match 0.8%; Score 18; DB 10; Length 479;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1160 CCAAGGCCCAACCGTGAGA 1177  
DB 380 CCAAGGCCCAACCGTGAGA 397

## RESULT 15

US-09-918-995-1375  
Sequence 1375, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1375  
LENGTH: 499  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(499)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-1375

Query Match 0.8%; Score 18; DB 9; Length 499;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 GACTATGCCATGACCATC 813  
DB 327 GACTATGCCATGACCATC 344

Search completed: May 23, 2003, 09:58:00  
Job time : 210 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 04:44:52 ; Search time 328 Seconds  
(without alignments)  
14686.050 Million cell updates/sec

Title: US-10-008-355-1  
Perfect score: 2139  
Sequence: 1 atgcaaatgaattataaag.....aagagctgaattgatctaa 2139

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_101002:\*\*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*\*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*\*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*\*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*\*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*\*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*\*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*\*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*\*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*\*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*\*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*\*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*\*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*\*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*\*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*\*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*\*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*\*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*\*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*\*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*\*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*\*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*\*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*\*
- 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2139	100.0	2139	24	AA143635
2	19	0.9	315	23	ABL27551
3	19	0.9	948	23	ABL03799
4	19	0.9	1431	24	ABQ70337
5	19	0.9	2252	23	ABL27548
6	19	0.9	2315	23	ABL27550
7	19	0.9	2324	23	ABL27552
8	19	0.9	2916	23	ABL03838
9	19	0.9	3006	23	ABL03798

10	19	0.9	4163	24	ABL65218	Lung cancer relate
11	19	0.9	4428	22	ABA09031	Human gap homolog
12	19	0.9	72928	20	AAZ18355	Human ASTH1J 5' ge
13	19	0.9	72928	21	AAA80253	Human ASTH1J 5' ge
14	19	0.9	1830121	17	AA142063	Haemophilus infliue
15	18	0.8	228	24	ABN69869	Streptococcus poly
16	18	0.8	252	24	ABK80475	Bacillus clausii g
17	18	0.8	287	16	AA123588	Human gene signatu
18	18	0.8	318	21	AA123588	Human foetal liver
19	18	0.8	466	22	ABA58492	Human foetal liver
20	18	0.8	466	22	ABA27556	Probe #6022 for ge
21	18	0.8	466	22	AAK06605	Human brain expres
22	18	0.8	466	22	AAK32297	Human bone marrow
23	18	0.8	466	22	AA116030	Probe #5963 for ge
24	18	0.8	466	22	AA138149	Probe #6835 used t
25	18	0.8	466	24	ABS07075	Human genome-deriv
26	18	0.8	577	21	AA154596	Arabidopsis thalia
27	18	0.8	579	21	AA153594	Arabidopsis thalia
28	18	0.8	550	24	ABQ59965	Human colon cancer
29	18	0.8	756	22	AAH04348	Human cDNA clone (
30	18	0.8	953	24	ABN74565	Bovine embryonic g
31	18	0.8	1032	23	ABL06001	Drosophila melanog
32	18	0.8	1251	18	AAV25022	H. pylori cell env
33	18	0.8	1379	23	ABL07075	Drosophila melanog
34	18	0.8	1446	22	AA151598	DNA encoding G-pro
35	18	0.8	1490	21	AA136486	Arabidopsis thalia
36	18	0.8	1491	21	AA136486	Arabidopsis thalia
37	18	0.8	1908	18	AAV25132	H. pylori cell env
38	18	0.8	1938	18	AAV24723	H. pylori ORF hplp
39	18	0.8	2007	18	AAV25118	H. pylori cell env
40	18	0.8	2070	21	AAZ88408	H. pylori bacteria
41	18	0.8	2114	21	AAZ87667	Human secreted pro
42	18	0.8	2115	20	AAZ75829	H. pylori outer me
43	18	0.8	2127	21	AAZ88405	H. pylori bacteria
44	18	0.8	2158	21	AAZ88409	H. pylori bacteria
45	18	0.8	2226	21	AAZ88406	H. pylori bacteria

ALIGNMENTS

RESULT 1  
AA143635  
ID AAL43635 standard; DNA; 2139 BP.  
XX  
AC AAL43635;  
XX  
DT 05-SEP-2002 (first entry)  
XX  
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.  
XX  
KW Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;  
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
KW periodontitis;  
XX  
OS Porphyromonas gingivalis.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2139  
FT /tag= a  
FT /product= "Porphyromonas gingivalis DPP-7"

PD	16-MAY-2002.
XX	
XX	08-NOV-2001; 2001WO-US46782.
XX	
XX	08-NOV-2000; 2000US-246827P.
XX	
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.
XX	
PI	Travis J, Potempa JS, Banbula A, Bugno M;

XX WPI: 2002-490075/52.  
DR P-PSDB: AAO15205.  
XX Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
PT the dipeptidylpeptidase for protecting an animal from periodontal  
PT disease caused by Porphyromonas gingivalis  
XX  
PS Claim 11; Fig 4; 65pp; English.  
XX  
CC The invention comprises the amino acid and coding sequence of the  
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (Dpp-7) enzyme. The Dpp-7  
CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
CC second and third amino acids from the N-terminal end of a target peptide.  
CC The Dpp-7 target peptide has an aliphatic or aromatic residue as a  
CC substituent on the alpha-carbon atom of the second amino acid from the  
CC N-terminal end. The Dpp-1 protein and DNA sequences of the invention are  
CC useful for identifying a Dpp-7 inhibitor. Dpp-7 inhibitors are useful for  
CC reducing the growth of a bacterium and protecting an animal from a  
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
CC or periodontitis). The present DNA sequence encodes the Porphyromonas  
CC gingivalis dipeptidylpeptidase-7 (Dpp-7) enzyme of the invention.  
XX  
SQ Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other;  
Query Match 100.0%; Score 2139; DB 24; Length 2139;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGCAATGAAATTAATAAGTATTCTTCGAGAGAGCCCTGCTGTGGGTGCTTCAGG 60  
Db 1 ATGCAATGAAATTAATAAGTATTCTTCGAGAGAGCCCTGCTGTGGGTGCTTCAGG 60  
Qy 61 GTACCAAGCGCAAGAGGATGTGCTCTCAAGNACTCAATCAGGAAATCTGGAT 120  
Db 61 GTACCAAGCGCAAGAGGATGTGCTCTCAAGNACTCAATCAGGAAATCTGGAT 120  
Qy 121 CGAATGCGTAGCTCGGCTTTACGCTCCCGTTGGATTGCGTCTACAGTTTCGACAAGCG 180  
Db 121 CGAATGCGTAGCTCGGCTTTACGCTCCCGTTGGATTGCGTCTACAGTTTCGACAAGCG 180  
Qy 181 TCCATTTGCAATGCGGTGTTATCTTCGGTGGCGGATGTACCGGTATCAGTGTCCGAT 240  
Db 181 TCCATTTGCAATGCGGTGTTATCTTCGGTGGCGGATGTACCGGTATCAGTGTCCGAT 240  
Qy 241 CAGGCGCTGATCTTTACCAACACACCTGCGGATACGCTGTATCCAGAGCAAGACG 300  
Db 241 CAGGCGCTGATCTTTACCAACACACCTGCGGATACGCTGTATCCAGAGCAAGACG 300  
Qy 301 GTGATCAGACTATCTGCGCGATGGTTTCGTTCTCGCAGATGGGTGAGAGCTTCCG 360  
Db 301 GTGATCAGACTATCTGCGCGATGGTTTCGTTCTCGCAGATGGGTGAGAGCTTCCG 360  
Qy 361 ATTCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAAAGGACAAAGTAA 420  
Db 361 ATTCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAAAGGACAAAGTAA 420  
Qy 421 GGACAGCTCAAGGGTATCAGTACGAGATGAGCGTCTGCGCAAGCTCAGGAGGTATGC 480  
Db 421 GGACAGCTCAAGGGTATCAGTACGAGATGAGCGTCTGCGCAAGCTCAGGAGGTATGC 480  
Qy 481 CAAGAAGCTGGCCAAAGAAATGACAGACGAGAACCAACTCTGCATCTGTAGAGCCTTTC 540  
Db 481 CAAGAAGCTGGCCAAAGAAATGACAGACGAGAACCAACTCTGCATCTGTAGAGCCTTTC 540  
Qy 541 TATTCACAAACGAATATCTCTCATCGTCTACGATGATTTCAAGGAGCTTCGTATGGTA 600  
Db 541 TATTCACAAACGAATATCTCTCATCGTCTACGATGATTTCAAGGAGCTTCGTATGGTA 600  
Qy 601 TTTGCTCTCCACGCTCTGTAGGTAAAGTTCGGAGCGGATACGGACAACCTGATGTGCCG 660  
Db 601 TTTGCTCTCCACGCTCTGTAGGTAAAGTTCGGAGCGGATACGGACAACCTGATGTGCCG 660

Qy 661 CGTCACACGGCGGACTTCAGCGTATTCGCGGTGTATGCGCGTCCGACAAACGCGCGGCC 720  
Db 661 CGTCACACGGCGGACTTCAGCGTATTCGCGGTGTATGCGCGTCCGACAAACGCGCGGCC 720  
Qy 721 GAATACAGCAAGGACAAATAAACCCCTATAAGCCCTTTTACTTCGCTCGGTATCCNTCAA 780  
Db 721 GAATACAGCAAGGACAAATAAACCCCTATAAGCCCTTTTACTTCGCTCGGTATCCNTCAA 780  
Qy 781 GGCTACAAAGCTGACGACTATGCCATGACCATCGGTTTCCGGGCGAGTACGGATCGCTAC 840  
Db 781 GGCTACAAAGCTGACGACTATGCCATGACCATCGGTTTCCGGGCGAGTACGGATCGCTAC 840  
Qy 841 CTCACCTTCTTGGGCTGTGGAAGATCGTATCGAAAACGAGAAACATCTCGTATCGAAGTT 900  
Db 841 CTCACCTTCTTGGGCTGTGGAAGATCGTATCGAAAACGAGAAACATCTCGTATCGAAGTT 900  
Qy 901 CGCGGTATCAAGCAAGGATCTGGAAGAGCCATGAGCGAGATCAGGCTACCCGATATC 960  
Db 901 CGCGGTATCAAGCAAGGATCTGGAAGAGCCATGAGCGAGATCAGGCTACCCGATATC 960  
Qy 961 AAATATGCCACCAAGTATGCTCAGAGTGCCTAACTATTGGAAGAAATTCGATCGGTATGAAC 1020  
Db 961 AAATATGCCACCAAGTATGCTCAGAGTGCCTAACTATTGGAAGAAATTCGATCGGTATGAAC 1020  
Qy 1021 CGCGGTCTCGCTCTTGTAGCTGATAGGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
Db 1021 CGCGGTCTCGCTCTTGTAGCTGATAGGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
Qy 1081 GACTGGATCCGTAAGAACGCGCAAGAGTGTCTATGCGGATGATTGCTCTCTCTCGAA 1140  
Db 1081 GACTGGATCCGTAAGAACGCGCAAGAGTGTCTATGCGGATGATTGCTCTCTCTCGAA 1140  
Qy 1141 AAGGCTTATAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200  
Db 1141 AAGGCTTATAGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200  
Qy 1201 TTCGGTGTACCGAGGTGTTGCTTTTGCACAGTTTGCACAGCTTGGCTACAAATCCT 1260  
Db 1201 TTCGGTGTACCGAGGTGTTGCTTTTGCACAGTTTGCACAGCTTGGCTACAAATCCT 1260  
Qy 1261 GATCCTCATGCGCGGTATCTCTCAAAATCGTTGACGACAAAGTACAAAGACTTACCTCCCTCG 1320  
Db 1261 GATCCTCATGCGCGGTATCTCTCAAAATCGTTGACGACAAAGTACAAAGACTTACCTCCCTCG 1320  
Qy 1321 CTCACCGTAAAGTGTGCGCGCATGCTCGATATGCTACGCGCGGTATGCTCTCGCGAC 1380  
Db 1321 CTCACCGTAAAGTGTGCGCGCATGCTCGATATGCTACGCGCGGTATGCTCTCGCGAC 1380  
Qy 1381 AAGCTCCCGGATATATCAAGAAATGTAATCGACAAGAAATTCAAAGGCGACACGAGAAG 1440  
Db 1381 AAGCTCCCGGATATATCAAGAAATGTAATCGACAAGAAATTCAAAGGCGACACGAGAAG 1440  
Qy 1441 TATCGACACTTCGTATTCGACAAAGAGTGTGGTTCCTTTATAGCGACAAGTTCCATGCCATG 1500  
Db 1441 TATCGACACTTCGTATTCGACAAAGAGTGTGGTTCCTTTATAGCGACAAGTTCCATGCCATG 1500  
Qy 1501 CTCAAGTCCATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560  
Db 1501 CTCAAGTCCATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560  
Qy 1561 CTTTCCAAAGAGCGTAAATAGCTGCTGCGGCTATTCAGGCGGATGCGATGGCCAAATGCC 1620  
Db 1561 CTTTCCAAAGAGCGTAAATAGCTGCTGCGGCTATTCAGGCGGATGCGATGGCCAAATGCC 1620  
Qy 1621 TATGCCAATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680  
Db 1621 TATGCCAATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680  
Qy 1681 CGTCTCTCGCGAGCGATGCCAACTTCACCATGCTATGAGTACGCTCCATCAAGGGA 1740  
Db 1681 CGTCTCTCGCGAGCGATGCCAACTTCACCATGCTATGAGTACGCTCCATCAAGGGA 1740  
Qy 1741 TATGAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800

Db 721 AACTGGCCAAAAAGAAA 739

Db 721 AACTGGCCAAAGAAAA 739



PI Venter JC, Adams M, Li PWD, Myers EW;  
DR WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 34123; 21pp + Sequence Listing; English.  
PS  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL01840-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2315 BP; 608 A; 443 C; 568 G; 696 T; 0 other;  
SQ

Query Match 0.9%; Score 19; DB 23; Length 2315;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 CACCACGCGGATACGGTG 280  
|||||  
Db 1002 CACCACGCGGATACGGTG 1020

RESULT 7  
ABL27552  
ID ABL27552 standard; DNA; 2324 BP.  
XX  
XX ABL27552;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 34129.  
DE  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
PN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
FR  
XX  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI; 2001-656860/75.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 34129; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2324 BP; 672 A; 414 C; 466 G; 772 T; 0 other;  
SQ

Query Match 0.9%; Score 19; DB 23; Length 2324;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 CACCACGCGGATACGGTG 280  
|||||  
Db 2068 CACCACGCGGATACGGTG 2086

RESULT 8  
ABL03838/c  
ID ABL03838 standard; cDNA; 2916 BP.  
XX  
XX ABL03838;  
AC  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 5996.  
DE  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
PN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
FR  
XX  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI; 2001-656860/75.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 5996; 21pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2916 BP; 739 A; 611 C; 476 G; 1090 T; 0 other;  
SQ

Query Match 0.9%; Score 19; DB 23; Length 2916;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



CC anti-neoplastic agent. The method involves exposing cells to a chemical  
 CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC result of M1, and the data is sufficient to convey the chemical  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilm's tumour.  
 CC  
 SQ Sequence 4163 BP; 836 A; 1303 C; 1112 G; 912 T; 0 other;

Query Match 0.9%; Score 19; DB 24; Length 4163;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 241 CAGGGCCTGATCTTTACCA 259

Db 3135 CAGGGCCTGATCTTTACCA 3153

# RESULT 11

IDA09031/C  
 ID ABA09031 standard; cDNA; 4428 BP.

XX AC ABA09031;

XX DT 11-JAN-2002 (first entry)

XX DE Human gap homologue-encoding cDNA, SEQ ID NO:807.

XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnery; antiulcer; ss.

XX OS Homo sapiens.

XX PN WO200157188-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US03800.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX XX WPI; 2001-457740/49.

DR P-PSDB; ABB11787.

XX Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -

XX Claim 1; Page 713-715; 1963pp; English.

XX Sequences ABA10981-ABBI2330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, thereby  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a cDNA encoding a  
 CC novel human polypeptide of the invention.

SQ Sequence 4428 BP; 968 A; 1198 C; 1372 G; 890 T; 0 other;

Query Match 0.9%; Score 19; DB 22; Length 4428;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 241 CAGGGCCTGATCTTTACCA 259

Db 1033 CAGGGCCTGATCTTTACCA 1015

# RESULT 12

AAZ18355/C

ID AAZ18355 standard; DNA; 72928 BP.

XX AC AAZ18355;

XX DT 19-OCT-1999 (first entry)

XX DE Human ASTH1J 5' genomic region.

XX KW ASTH1; asthma; human; chromosome 11p; ASTH1J; ASTH1J; genetic locus;  
 KW therapeutic; immunogen; ds.

XX OS Homo sapiens.

XX PN WO9937809-A1.

XX PD 29-JUL-1999.

XX XX 21-JAN-1998; 98WO-US01260.

XX PF



```

PR 21-JAN-1998; 98WO-US01260.
XX (AXYS-) AXYS PHARM INC.
PI Brooks-Wilson AR, Buckler A, Cardon L, Carey AH;
PI Galvin M, Miller A, North M;
XX WPI; 1999-479058/40.
XX Mammalian asthma related genes, useful for diagnosis of a
PT predisposition to development of asthma
XX Claim 15; Page 75-96; 195pp; English.
XX The invention identifies a genetic locus ASTH1, associated with asthma,
CC mapped to human chromosome 11p. ASTH1I and ASTH1J are genes present
CC within the locus, located close to each other on human chromosome 11p,
CC and have similar patterns of expression, and common sequence motifs. The
CC ASTH1 genes and fragments, encoded protein, genomic regulatory regions
CC and anti-ASTH1 antibodies are useful in the identification of individuals
CC predisposed to development of asthma, and for the modulation of gene
CC activity in vivo for prophylactic and therapeutic purposes. The ASTH1
CC protein is useful as an immunogen to raise specific antibodies, in drug
CC screening for compositions that mimic or modulate ASTH1 activity or
CC expression, including altered forms of ASTH1 protein, and as a
CC therapeutic. The present sequence represents a human ASTH1 genomic region
CC sequence.
XX
SQ Sequence 72928 BP; 20042 A; 15100 C; 16036 G; 21750 T; 0 other;

Query Match 0.9%; Score 19; DB 20; Length 72928;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1625 CCATTGAGAGGCGCAAGCG 1643
DB 42822 CCATTGAGAGGCGCAAGCG 42804

RESULT 13
AAA80253/C
ID AAA80253 standard; DNA; 72928 BP.
XX
AC AAA80253;
XX
DT 22-NOV-2000 (first entry)
XX
DE Human ASTH1J 5' genomic region.
XX
KW ASTH1 locus; ASTH1J; human; chromosome 11p; asthma;
KW bronchial hyperreactivity; ets family; transcription factor;
KW splice variant; genetic predisposition; polymorphism; antibody;
KW drug screening; prophylaxis; therapy; diagnosis; ds.
XX
OS Homo sapiens.
XX
XX US6087485-A.
XX
XX 11-JUL-2000.
XX
XX 21-JAN-1998; 98US-0009913.
XX
XX 21-JAN-1997; 97US-0035663.
XX
XX 01-JUL-1997; 97US-0051432.
XX
XX (AXYS-) AXYS PHARM INC.
XX
XX Galvin M, Miller A, North M, Cardon L, Buckler A;
XX Brooks-Wilson AR, Carey AH;
XX WPI; 2000-505109/45.
XX
XX New nucleic acids other than naturally occurring chromosomes encoding
PT

```

---

```

PT ASTH1 protein, for e.g. screening compositions that modulate expression
PT or function of ASTH1 proteins or as diagnostics for genetic
PT predisposition to asthma
XX
XX Claim 7; Column 49-112; 131pp; English.
XX
XX The invention relates to the ASTH1 locus on the short arm of human
CC chromosome (11p). This locus comprises the ASTH1I and ASTH1J genes,
CC which are associated with a genetic predisposition to asthma and
CC bronchial hyperreactivity. The ASTH1I and ASTH1J genes are oriented in
CC opposite directions with the ASTH1 locus, and have similar patterns of
CC expression and common sequence motifs. They are both expressed in
CC trachea, lung and several other tissues. ASTH1I and ASTH1J are novel
CC members of the ets family of transcription factors, which have been
CC implicated in the activation of a variety of genes including the TCRA
CC gene and cytokine genes known to be important in the aetiology of asthma.
CC Both ASTH1I and ASTH1J mRNAs are alternatively spliced. Alternative
CC splicing of transcripts has no effect on the open reading frame of
CC ASTH1J, as the exons involved are all 5' to the start codon in exon b. In
CC contrast, alternative splicing of ASTH1I transcripts results in 3
CC different ASTH1 isoforms. The invention also encompasses mouse asth1j
CC protein. The ASTH1 nucleic acids are useful as diagnostics to identify a
CC hereditary predisposition to asthma, as probes for identifying ASTH1
CC related genes, for identifying expression of the gene in a biological
CC specimen, and for generating genetically modified non-human animals or
CC site specific gene modifications in cell lines. The encoded ASTH1
CC proteins are useful as immunogens to raise specific antibodies; in drug
CC screening for compositions that mimic or modulate activity or expression
CC of ASTH1I and/or ASTH1J (including altered forms of these proteins); and
CC as a therapeutic. The ASTH1 genes or fragments thereof, encoded proteins,
CC ASTH1 genomic regulatory regions, and anti-ASTH1I and anti-ASTH1J
CC antibodies are useful in the identification of individuals predisposed to
CC development of asthma, and for modulation of gene activity in vivo for
CC prophylactic and therapeutic purposes. The intact ASTH1I or ASTH1J
CC proteins or active fragments thereof may be used to modulate or reduce
CC bronchial hyperreactivity. The present sequence represents the
CC non-transcribed region upstream of the human ASTH1J gene.
XX
SQ Sequence 72928 BP; 20041 A; 15101 C; 16036 G; 21750 T; 0 other;

Query Match 0.9%; Score 19; DB 21; Length 72928;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1625 CCATTGAGAGGCGCAAGCG 1643
DB 42822 CCATTGAGAGGCGCAAGCG 42804

RESULT 14
AAT42063/C
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
XX 14-SEP-1999 (first entry)
XX
XX Haemophilus influenzae complete genome sequence.
XX
XX Genome: bacterium; Haemophilus influenzae; computer readable medium;
XX expression modulating fragment; regulation; gene expression; vector;
XX organism; open reading frame; ORF; ds.
XX
XX Haemophilus influenzae.
XX
XX WO9633276-A1.
XX
XX 24-OCT-1996.
XX
XX 22-APR-1996; 96WO-US05320.
XX
XX 07-JUN-1995; 95US-0487429.
XX
XX 21-APR-1995; 95US-0426787.
XX

```

```

PR 07-JUN-1995; 95US-0476102.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UVOJ ) UNIV JOHNS HOPKINS.
XX
XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX WPI; 1996-485782/48.
XX
XX Haemophilus influenzae Rd genome recorded on computer readable
XX medium - useful for identifying commercially important nucleic acid
XX fragments by homology searching
XX
XX Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
XX This sequence represents the complete genome sequence of the bacterium
XX Haemophilus influenzae strain Rd. The invention relates to a computer
XX readable medium (CRM) having recorded upon it the complete H.influenzae
XX nucleotide sequence (I), a representative fragment of (I) or a nucleotide
XX sequence at least 99% identical to (I). By providing the full-length
XX genomic sequence in a computer readable form, it is possible to identify
XX commercially important nucleic acid fragments and expression modulating
XX fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
XX regulate the expression of a nucleic acid molecule. Vectors and altered
XX organisms comprising the predicted ORFs can be used to produce any of the
XX polypeptide fragments of the H. influenzae Rd genome.
XX
XX Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;
SQ
Query Match 0.9%; Score 19; DB 17; Length 1830121;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1138 GAAAAGGCTTATAGGAAG 1156
|||||
Db 1428299 GAAAAGGCTTATAGGAAG 1428281

RESULT 15
ABN69869
ID ABN69869 standard; DNA; 228 BP.
XX
XX AC ABN69869;
XX
XX 01-JUL-2002 (first entry)
XX
XX Streptococcus polynucleotide SEQ ID NO 7651.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
XX Streptococcus pyogenes.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
XX 24-NOV-2000; 2000GB-0028727.
XX 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
XX Tettelin H;
XX
XX WPI; 2002-352536/38.
XX P-PSDB; ABP29236.
XX

```

---

```

PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein.
XX
XX Claim 7; Page 3903; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX antibodies that bind (I) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (I) are used to detect Streptococcus in a
XX biological sample. (I) is used to determine whether a compound binds to
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (I) may be used to recombinantly produce (I) and may be
XX used in gene therapy. Antibodies to (I) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.
XX
XX Sequence 228 BP; 85 A; 23 C; 46 G; 74 T; 0 other;
SQ
Query Match 0.8%; Score 18; DB 24; Length 228;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 ATGAATTTAAAAAGTATT 24
|||||
Db 4 ATGAATTTAAAAAGTATT 21

Search completed: May 23, 2003, 07:31:39
Job time : 2006 secs

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 05:20:28 ; Search time 62 Seconds  
(without alignments)  
10580.348 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaaatgaattataaaag.....aagagctgaattgatctaa 2139

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 15338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	1500	70.1	1974	4	US-09-221-017B-726
2	92	4.3	561	4	US-09-221-017B-12
3	19	0.9	72928	3	US-09-009-913-1
4	18	0.8	1317	4	US-09-221-017B-382
5	18	0.8	4403765	4	US-09-103-840A-2
6	18	0.8	4411529	4	US-09-103-840A-1
7	17	0.8	861	2	US-08-743-637B-166
8	17	0.8	861	3	US-08-526-840B-166
9	17	0.8	1958	4	US-09-570-454-1
10	17	0.8	2352	4	US-08-997-251-3
11	17	0.8	2692	1	US-07-932-454A-2
12	17	0.8	2822	1	US-08-679-405-1
13	17	0.8	2822	2	US-08-842-799-1
14	17	0.8	2822	5	PCT/US96-11458-1
15	17	0.8	49272	1	US-08-614-770A-1
16	16	0.7	50	2	US-08-832-468-5
17	16	0.7	313	1	US-08-396-452-1
18	16	0.7	313	4	US-09-169-119-1
19	16	0.7	360	4	US-09-060-756-211
20	16	0.7	592	4	US-09-385-982-126
21	16	0.7	615	4	US-08-998-416-1063
22	16	0.7	652	4	US-09-328-111-161
23	16	0.7	705	4	US-08-998-416-1059
24	16	0.7	735	4	US-09-535-008-32
25	16	0.7	1063	4	US-09-077-675A-1
26	16	0.7	1322	4	US-09-446-821A-2
27	16	0.7	1330	1	US-08-480-604A-22

ALIGNMENTS

RESULT 1

US-09-221-017B-726  
; Sequence 726, Application US/09221017B  
; Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 726:

SEQUENCE CHARACTERISTICS:

LENGTH: 1974 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

28 0.7 1330 2 US-08-405-496A-22 Sequence 22, Appl  
29 16 0.7 1330 4 US-08-915-136-22 Sequence 22, Appl  
30 16 0.7 1347 2 US-08-959-749-1 Sequence 1, Appl  
31 16 0.7 1347 4 US-09-351-497-1 Sequence 1, Appl  
32 16 0.7 1359 4 US-09-134-001C-1336 Sequence 1336, Ap  
33 16 0.7 1402 1 US-08-480-604A-25 Sequence 25, Appl  
34 16 0.7 1402 2 US-08-405-496A-25 Sequence 25, Appl  
35 16 0.7 1402 4 US-08-915-136-25 Sequence 25, Appl  
36 16 0.7 1491 4 US-08-740-223A-21 Sequence 21, Appl  
37 16 0.7 1491 4 US-09-709-188-21 Sequence 21, Appl  
38 16 0.7 1500 4 US-08-740-223A-23 Sequence 23, Appl  
39 16 0.7 1500 4 US-09-709-188-23 Sequence 23, Appl  
40 16 0.7 1540 1 US-08-286-872-5 Sequence 5, Appl  
41 16 0.7 1812 1 US-08-328-322-18 Sequence 18, Appl  
42 16 0.7 1829 1 US-08-201-118-8 Sequence 8, Appl  
43 16 0.7 1829 2 US-08-238-821B-8 Sequence 8, Appl  
44 16 0.7 1829 5 PCT-US95-05744-8 Sequence 8, Appl  
45 16 0.7 1964 1 US-08-328-322-9 Sequence 9, Appl

P  
D

198  
102e  
1500

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORYPHYROMONAS GINGIVALIS  
FEATURE:

NAME/KEY: misc\_feature  
LOCATION: 1...1974

US-09-221-017B-726

Query Match 70.1%; Score 1500; DB 4; Length 1974;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1730; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 409 GACAAAGGTAGAGGACAGCTCAAGGGTATCACTACGAGATGGAGCGTCTCGCAAGAGCT 468  
DB 1 GACAAAGGTAGAGGACAGCTCAAGGGTATCACTACGAGATGGAGCGTCTCGCAAGAGCT 60  
QY 469 CAGGAGGTATGCCAAGAACTGGCCAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 528  
DB 61 CAGGAGGTATGCCAAGAACTGGCCAAAAGAAAATGCAGACGAGAACCAACTCTGCATC 120  
QY 529 GTAGAGCCTTTCTATTCCAACAAAGTAATCTCTCATCTGCTACGATGATTCAAGAC 588  
DB 121 GTAGAGCCTTTCTATTCCAACAAAGTAATCTCTCATCTGCTACGATGATTCAAGAC 180  
QY 589 GTTCGTATGGTATTTGCTCTCCAGCTCTCTAGGTAAAGTTTCGGAGGCGATACGAGAAC 648  
DB 181 GTTCGTATGGTATTTGCTCTCCAGCTCTCTAGGTAAAGTTTCGGAGGCGATACGAGAAC 240  
QY 649 TGGATGTGGCCGCTCACACGGGCGACTTTCAGCGTATTCGCGGTGATCGCGGTGCCGAC 708  
DB 241 TGGATGTGGCCGCTCACACGGGCGACTTTCAGCGTATTCGCGGTGATCGCGGTGCCGAC 300  
QY 709 AACCGCGCGCGGTAATACAGCAAGCAATAAACCTTAAAGCCGTTTACTTCGCTGCC 768  
DB 301 AACCGCGCGCGGTAATACAGCAAGCAATAAACCTTAAAGCCGTTTACTTCGCTGCC 360  
QY 769 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 828  
DB 361 GTATCCATGCAAGGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 420  
QY 829 ACGGATCGCTACCTACCTTCTTTGGGGTGTGGAAGATCGTATCGAAAAACGAGAACATCCT 888  
DB 421 ACGGATCGCTACCTACCTTCTTTGGGGTGTGGAAGATCGTATCGAAAAACGAGAACATCCT 480  
QY 889 CGTATCGAAGTTCCGGTATCAACGAGGATCTGGAAGGAGCCATGACGCGAGATCAG 948  
DB 481 CGTATCGAAGTTCCGGTATCAACGAGGATCTGGAAGGAGCCATGACGCGAGATCAG 540  
QY 949 GCTACCGGTATCAAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAAATTCG 1008  
DB 541 GCTACCGGTATCAAAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTGGAAGAAATTCG 600  
QY 1009 ATCGGTATGAACCCGGTCTCGCTGCTTTGACGTGATAGGTGTAAGCGTGGCGAGGAA 1068  
DB 601 ATCGGTATGAACCCGGTCTCGCTGCTTTGACGTGATAGGTGTAAGCGTGGCGAGGAA 660  
QY 1069 AGACATTCGAGACTGGATCCGTAAGAACGGAAGAGTGTGCTATGGCGATGATTG 1128  
DB 661 AGACATTCGAGACTGGATCCGTAAGAACGGAAGAGTGTGCTATGGCGATGATTG 720  
QY 1129 TCTTCTCTCGAAAAGGCTTATAAGGAAGGAGGAGCCAAAGCCACCGTGAGATGACTTATTG 1188  
DB 721 TCTTCTCTCGAAAAGGCTTATAAGGAAGGAGGAGCCAAAGCCACCGTGAGATGACTTATTG 780  
QY 1189 AGCAGAGCGCTTTCGGTGGTACCGAGGTGGTTCGTTTGCACAGTTTGCACAGCATATG 1248  
DB 781 AGCAGAGCGCTTTCGGTGGTACCGAGGTGGTTCGTTTGCACAGTTTGCACAGCATATG 840  
QY 1249 GCTACAATCTGTATGCTATGCGGTATCCTCAAAATCGTTGACGACAACTACAAAGAC 1308  
DB 841 GCTACAATCTGTATGCTATGCGGTATCCTCAAAATCGTTGACGACAACTACAAAGAC 900

QY 1309 TACCTCCCTCGCTCGACCGTAAGTGTCTGCCGCCATGCTCGATATTGTACGCCGCGT 1368  
DB 901 TACCTCCCTCGCTCGACCGTAAGTGTCTGCCGCCATGCTCGATATTGTACGCCGCGT 960  
QY 1369 ATCCCTGCCGACAAAGCTCCCGGATATATTCAAGAAATGTAATCGACAAAGAAATTCAAAGGC 1428  
DB 961 ATCCCTGCCGACAAAGCTCCCGGATATATTCAAGAAATGTAATCGACAAAGAAATTCAAAGGC 1020  
QY 1429 GACACGAAGAGTATGCAGACTTCGTATTCGACAAAGAGTGTGTTCTCTTATAGGACAAAG 1488  
DB 1021 GACACGAAGAGTATGCAGACTTCGTATTCGACAAAGAGTGTGTTCTCTTATAGGACAAAG 1080  
QY 1489 TTCCATGCCATGCTCAAGTCCATGCACAAAGAAAGTTTGCACAAAGGCTATTCGACAAAGAT 1548  
DB 1081 TTCCATGCCATGCTCAAGTCCATGCACAAAGAAAGTTTGCACAAAGGCTATTCGACAAAGAT 1140  
QY 1549 CCGCAGTAGAGCTTTCCAAAGAGCGTAAATAGTCTGCTCGCGCTATTCAGGCGCGATGCG 1608  
DB 1141 CCGCAGTAGAGCTTTCCAAAGAGCGTAAATAGTCTGCTCGCGCTATTCAGGCGCGATGCG 1200  
QY 1609 ATGCCCAATGCCATTCGCAATTCGACAAAGGCGCAAGCGTCTTTTCTTTGCGCGTTTCGCTGAG 1668  
DB 1201 ATGCCCAATGCCATTCGCAATTCGACAAAGGCGCAAGCGTCTTTTCTTTGCGCGTTTCGCTGAG 1260  
QY 1669 ATGTACCCCGGACGCTGCTCTGCCGAGCGATGCCAACTTTCACCATCGCTATGAGCTACGCG 1728  
DB 1261 ATGTACCCCGGACGCTGCTCTGCCGAGCGATGCCAACTTTCACCATCGCTATGAGCTACGCG 1320  
QY 1729 TCCATCAAGGGATATGAACCGGACGCGTGCCTGTACAACTATATACGACGAGCAAG 1788  
DB 1321 TCCATCAAGGGATATGAACCGGACGCGTGCCTGTACAACTATATACGACGAGCAAG 1380  
QY 1789 GGCCTATTCGAGAGCAGGAT--CCTAAGAGCGATGAGTTTGCCTACAGGAGAAATATCC 1846  
DB 1381 GGCCTATTCGAGAGCAGGATTCCTTAAGAGCGATGAGTTTGCCTACAGGAGAAATATCC 1440  
QY 1847 TCGACCTCTTCCGACCAAACTATGCTCGCTATGCGGAGAACGCTGAGCTCCATATCG 1906  
DB 1441 TCGACCTCTTCCGACCAAACTATGCTCGCTATGCGGAGAACGCTGAGCTCCATATCG 1500  
QY 1907 CTTTCTCTCGAACAAGACATACGCGGCGGTAACCTCGGTAGCCCGCTATTCGATAGA 1966  
DB 1501 CTTTCTCTCGAACAAGACATACGCGGCGGTAACCTCGGTAGCCCGCTATTCGATAGA 1560  
QY 1967 ACGCCGCTGATCGGCTTTCGCTTCGATGCAACTGCGAAGCTATGAGTGGTACATCG 2026  
DB 1561 ACGCCGCTGATCGGCTTTCGCTTCGATGCAACTGCGAAGCTATGAGTGGTACATCG 1620  
QY 2027 AGTTTCAACCCGATCTCGACGCAATACGCGTGGACATCCGCTAGCTTCTCTTCATGA 2086  
DB 1621 AGTTTCAACCCGATCTCGACGCAATACGCGTGGACATCCGCTAGCTTCTCTTCATGA 1680  
QY 2087 TTGACAAATGGGTGAGTGGCCCGCTCTCATCCAAAGAGCTGAAGTTGATCTAA 2139  
DB 1681 TTGACAAATGGGTGAGTGGCCCGCTCTCATCCAAAGAGCTGAAGTTGATCTAA 1733

## RESULT 2

US-09-221-017B-12  
; Sequence 12, Application US/09221017B  
; Patent No. 6444799  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...561
; US-09-221-017B-12

Query Match 4.3%; Score 92; DB 4; Length 561;
Best Local Similarity 99.3%; Pred. No. 1e-38;
Matches 142; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAATGAATTAATAAGTATTCTTCGGAGCAGCCCTGCTGTGGTGTCTTCAGGG 60
|
Db 366 ATGCAATGAATTAATAAGTATTCTTCGGAGCAGCCCTGCTGTGGTGTCTTAAGG 425
|

QY 61 GTAGCCAAAGCCGACAAAGCATGTGCTCTCAAGCACTCAATCAGGAGATCTGGAT 120
|
Db 426 GTAGCCAAAGCCGACAAAGCATGTGCTCTCAAGCACTCAATCAGGAGATCTGGAT 485
|

QY 121 CGAATGCGTGAGCTCGCTTAC 143
|
Db 486 CGAATGCGTGAGCTCGCTTAC 508
|

RESULT 3
US-09-009-913-1/c
; Sequence 1, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: AxyS Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
```

```
;
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,913
; FILING DATE: 21-JAN-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SEQ-4P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3231
; TELEFAX: 650-327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; US-09-009-913-1

Query Match 0.9%; Score 19; DB 3; Length 72928;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1625 CCATTGAGAGGCGCAAGCG 1643
|
Db 42822 CCATTGAGAGGCGCAAGCG 42804
|

RESULT 4
US-09-221-017B-382
; Sequence 382, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
```

APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA: PCT/AU98/01023  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: MONROY, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELEPHONE: 850-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 382:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1317 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORYPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1...1317  
US-09-221-017B-382

Query Match 0.8%; Score 18; DB 4; Length 1317;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 636 CGATACGACAACTGGAT 653  
Db 549 CGATACGACAACTGGAT 566

RESULT 5  
US-09-103-840A-2  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 0.8%; Score 18; DB 4; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 ACCGTAAGTGCTGCCG 1342  
Db 735985 ACCGTAAGTGCTGCCG 736002

RESULT 6  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 0.8%; Score 18; DB 4; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1325 ACCGTAAGTGCTGCCG 1342  
Db 734553 ACCGTAAGTGCTGCCG 734570

RESULT 7  
US-08-743-637B-166  
Sequence 166, Application US/08743637B  
Patent No. 5994066  
GENERAL INFORMATION:  
APPLICANT: BERGERON, Michel G.  
APPLICANT: PICARD, Francois J.  
APPLICANT: OUELLETTE, Marc  
APPLICANT: ROY, Paul H.  
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA  
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED  
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: QUARLES & BRADY  
STREET: 411 EAST WISCONSIN AVENUE  
CITY: MILWAUKEE  
STATE: WISCONSIN  
COUNTRY: USA  
ZIP: 53202-4497  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/743,637B  
FILING DATE: 04-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/526,840  
FILING DATE: 11-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: BAKER, Jean C.  
REGISTRATION NUMBER: 35,433  
REFERENCE/DOCKET NUMBER: 850586.90012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5000  
TELEFAX: (414) 277-5591  
INFORMATION FOR SEQ ID NO: 166:



```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-743-637B-166

Query Match          0.8%; Score 17; DB 2; Length 861;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 GCCTGCTGTTGGTGTC 53
Db 466 GCCCTGCTGTTGGTGTC 482

RESULT 8
; Sequence 166, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526.840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 166:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 861 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-526-840B-166

Query Match          0.8%; Score 17; DB 3; Length 861;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 GCCTGCTGTTGGTGTC 53
Db 466 GCCCTGCTGTTGGTGTC 482
```

```
RESULT 9
US-09-570-454-1/c
; Sequence 1, Application US/09570454
; Patent No. 6399743
; GENERAL INFORMATION:
; APPLICANT: Department of Veterans Affairs
; TITLE OF INVENTION: Isolation and characterization of epidermal growth
; FILE REFERENCE: 107999.00106
; CURRENT APPLICATION NUMBER: US/09/570,454
; CURRENT FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 60/134,200
; PRIOR FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1958
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-570-454-1

Query Match          0.8%; Score 17; DB 4; Length 1958;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1613 CCAATGCCTATGCCATT 1629
Db 1255 CCAATGCCTATGCCATT 1239

RESULT 10
US-08-997-251-3/c
; Sequence 3, Application US/08997251
; Patent No. 6271440
; GENERAL INFORMATION:
; APPLICANT: GUBLER, FRANZ J.
; APPLICANT: JACOBSEN, JOHN V.
; TITLE OF INVENTION: PLANT REGULATORY PROTEINS III
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Boulevard
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,251
; FILING DATE: 23-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO AU96/00383
; FILING DATE: 21-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PN6470/95
; FILING DATE: 09-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PN3779/95
; FILING DATE: 23-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Winner, Ellen P.
; REGISTRATION NUMBER: 28,547
; REFERENCE/DOCKET NUMBER: 110-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 3:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2352 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 396..2054
; US-08-997-251-3

Query Match      0.8%; Score 17; DB 4; Length 2352;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 601 TTGTGCTCTCCCGAGCTC 617
Db 309 TTGTGCTCTCCCGAGCTC 293

RESULT 11
US-07-932-454A-2/c
; Sequence 2, Application US/07932454A
; Patent No. 5262318
; GENERAL INFORMATION:
; APPLICANT: GUTHRIE, ELLEN P.
; TITLE OF INVENTION: ISOLATED DNA ENCODING THE SphI
; TITLE OF INVENTION: RESTRICTION ENDONUCLEASE AND RELATED METHODS FOR PRODUCING
; TITLE OF INVENTION: THE SAME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/932.454A
; FILING DATE: 19920820
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: 42078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2692 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 703..1653
; OTHER INFORMATION: /note= "METHYLASE GENE STARTS AT
; OTHER INFORMATION: POSITION 703/ENDS AT 1653. "RESTRICTION
; OTHER INFORMATION: ENDONUCLEASE STARTS AT POSITION 1703/ENDS AT 2410"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1703..2410
; US-07-932-454A-2
```

```
Query Match      0.8%; Score 17; DB 1; Length 2692;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1394 TATTCAGAATGTAATC 1410
Db 1809 TATTCAGAATGTAATC 1793

RESULT 12
US-08-679-405-1
; Sequence 1, Application US/08679405
; Patent No. 5866393
; GENERAL INFORMATION:
; APPLICANT: Fuglsang, Claus
; APPLICANT: Halkier, Torben
; APPLICANT: Oxenboll, Karen M.
; APPLICANT: Berka, Randy M.
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: Haloperoxidases from Curvularia
; TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58663930 No. 5866393disk of No. 5866393th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/679,405
; FILING DATE: July 9, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,194
; FILING DATE: July 14, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/603,534
; FILING DATE: February 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4441.210.US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 477..2276
; US-08-679-405-1

Query Match      0.8%; Score 17; DB 2; Length 2822;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GCAAAATGAAATTAATAA 19
Db 2629 GCAAAATGAAATTAATAA 2645
```

```

RESULT 13
US-08-842-799-1
; Sequence 1, Application US/08842799
; Patent No. 5965418
; GENERAL INFORMATION:
; APPLICANT: Fuglsang, Claus
; APPLICANT: Halkier, Torben
; APPLICANT: Oxenboll, Karen M.
; APPLICANT: Berka, Randy M.
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: Haloperoxidases from Curvularia
; TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59654180 No. 5965418th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/842,799
; FILING DATE: 16-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,405
; FILING DATE: July 9, 1996
; APPLICATION NUMBER: 60/001,194
; FILING DATE: July 14, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/603,534
; FILING DATE: February 21, 1996
; NAME: Lambiris, Elias J.
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4441.210.US
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 477..2276
US-08-842-799-1

Query Match 0.8%; Score 17; DB 2; Length 2822;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCAAAATGAATTAATAAAA 19
Db 2629 GCAAAATGAATTAATAAAA 2645

RESULT 14
PCT-US96-11458-1
; Sequence 1, Application PC/TUS9611458
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Haloperoxidases from Curvularia
; TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59654180 No. 5965418th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/842,799
; FILING DATE: 16-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,405
; FILING DATE: July 9, 1996
; APPLICATION NUMBER: 60/001,194
; FILING DATE: July 14, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/603,534
; FILING DATE: February 21, 1996
; NAME: Lambiris, Elias J.
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4441.210.US
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 477..2276
US-08-842-799-1

Query Match 0.8%; Score 17; DB 2; Length 2822;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCAAAATGAATTAATAAAA 19
Db 2629 GCAAAATGAATTAATAAAA 2645

RESULT 15
US-08-614-770A-1
; Sequence 1, Application US/08614770A
; Patent No. 5773267
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
; TITLE OF INVENTION: D29 SHUTTLE PHASIDS AND USES THEREOF
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:

```

```

; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/11458
; FILING DATE: 9-JUL-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,194
; FILING DATE: 14-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/603,534
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4441.204-WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 867-0123
; TELEFAX: (212) 878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2822 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 477..2276
PCT-US96-11458-1

Query Match 0.8%; Score 17; DB 5; Length 2822;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCAAAATGAATTAATAAAA 19
Db 2629 GCAAAATGAATTAATAAAA 2645

RESULT 15
US-08-614-770A-1
; Sequence 1, Application US/08614770A
; Patent No. 5773267
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
; TITLE OF INVENTION: D29 SHUTTLE PHASIDS AND USES THEREOF
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
; MEDIUM TYPE: DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/614,770A  
FILING DATE: MARCH 7, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: ELIZABETH A. BOGOSIAN  
REGISTRATION NUMBER: 39,911  
REFERENCE/DOCKET NUMBER: 96700/402  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 49272  
TYPE: NUCLEIC ACID  
STRANDEDNESS: DOUBLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE:  
DESCRIPTION: OLIGONUCLEOTIDE  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: MYCOBACTERIOPHAGE  
INDIVIDUAL ISOLATE: D29  
US-08-614-770A-1

Query Match 0.8%; Score 17; DB 1; Length 49272;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 836 GCTACCTCACTTCTTGG 852  
|  
Db 46900 GCTACCTCACTTCTTGG 46916

Search completed: May 23, 2003, 10:27:15  
Job time : 6285 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM' nucleic - nucleic search, using sw model

Run on: May 23, 2003, 05:02:37 ; Search time 2048 Seconds  
(without alignments)  
16915.110 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaaatgaattaaaag.....aagagctgaagtgtatctaa 2139

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpi:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	21	1.0	444	17	AQ820052 HS_5270_A
2	21	1.0	591	9	A1928924 au58a07.y
3	21	1.0	800	17	AG084100 Pan trogl
4	21	1.0	1061	13	BM553311 AGENCOURT
5	20	0.9	223	10	AW200173 dai2e08.y
6	20	0.9	257	10	AW881153 QVO-OT003

c 7	20	0.9	260	12	BG158262
c 8	20	0.9	410	10	AW199497
c 9	20	0.9	419	12	BE809917
c 10	20	0.9	463	14	BQ412487
c 11	20	0.9	472	14	BQ403470
c 12	20	0.9	473	12	BF654963
c 13	20	0.9	509	9	A1718392
c 14	20	0.9	519	12	BE809904
c 15	20	0.9	523	13	BI847597
c 16	20	0.9	531	13	BJ080132
c 17	20	0.9	547	10	AW653212
c 18	20	0.9	568	17	CNS07HUQ
c 19	20	0.9	570	9	AA582263
c 20	20	0.9	668	13	BJ043920
c 21	20	0.9	787	13	BM029390
c 22	20	0.9	988	13	BM321169
c 23	19	0.9	313	14	D68316
c 24	19	0.9	334	10	AW597013
c 25	19	0.9	352	17	AQ057720
c 26	19	0.9	358	14	T68361
c 27	19	0.9	360	14	D68463
c 28	19	0.9	360	14	D69805
c 29	19	0.9	368	14	BQ301839
c 30	19	0.9	368	14	BQ301843
c 31	19	0.9	369	17	BH07962
c 32	19	0.9	376	10	BB09399
c 33	19	0.9	402	9	AA062801
c 34	19	0.9	414	17	AZ841761
c 35	19	0.9	418	14	R50228
c 36	19	0.9	448	14	H73815
c 37	19	0.9	462	14	T71455
c 38	19	0.9	498	13	BJ107698
c 39	19	0.9	510	17	BH412712
c 40	19	0.9	522	17	AQ791029
c 41	19	0.9	523	17	AQ844615
c 42	19	0.9	524	17	BH01026
c 43	19	0.9	565	9	AI606604
c 44	19	0.9	587	13	BM504607
c 45	19	0.9	599	9	AU205200

ALIGNMENTS

RESULT 1  
AQ820052  
LOCUS  
DEFINITION  
HS\_5270\_A1\_H10\_SP6E RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=846 Col=19 Row=O, DNA sequence.  
ACCESSION  
AQ820052  
VERSION  
AQ820052.1  
KEYWORDS  
GSS.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 444)  
AUTHORS  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE  
99380589  
COMMENT  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong

AQ820052 444 bp DNA linear GSS 26-AUG-1999  
HS\_5270\_A1\_H10\_SP6E RPCI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate=846 Col=19 Row=O, DNA sequence.

AQ820052.1 GI:5782445

GSS.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 444)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

Site\_1: SstI; Site\_2: XhoI; Double-stranded cDNA was prepared from human fetal brain tissue. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence:  
5'-GAGAGAGAGAGAGCTCAGGATCTTAATTAAATTATCCCCCCCCCCCCC-3' and 3' adaptor sequence:  
5'-GAGAGAGAGAGCTCGATCTTTTTTTTTTTTTTT-3'. The library was size-selected for >0.5 kb inserts and has an average insert size estimated at 1.2 kb. This library was constructed using the CAP-trapper method for full-length enrichment and has not undergone amplification. Library was constructed by Dr. Claudio Schneider (LNCIB-Atea Science Park, Trieste, Italy).

enrichment and has not undergone amplification. The library was constructed by Dr. Claudio Schneider (LNCIB-Area Science Park, Trieste, Italy)."

BASE COUNT	115 a	207 c	164 g	104 t	1 others
ORIGIN					

Query Match 1.0%; Score 21; DB 9; Length 591;  
Best Local Similarity 100.0%; Pred. No. 21;  
ORIGIN

Qy	1160	CCAAGGCCAACCGTGAGATGA	1180
Dδ	409	CCAAGGCCAACCGTGAGATGA	429

DD 409 CCAGGCCAACCGGAGAGGA 429

RESULT 3

AG084100	AG084100	800 bp	DNA	linear	GSS 03-NOV-2001
LOCUS	Pan troglodytes DNA, clone: PTB-081L22.R,				genomic survey sequence.
DEFINITION	AG084100				
ACCESSION	AG084100.1				
VERSION	GI:16635902				

**SOURCE** Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male BAC Library clone:PTB-081L22.R.

ORGANISM  
pan troglodytes  
BAC Library clone:PTB-US1122.1.K.  
Phylum: Metazoa; Chordata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Placentalia; Primates; Catarrhini; Hominoidea; Pan.

REFERENCE	1
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	BAC end sequences of Library PTB
JOURNAL	Unpublished
PROCESSED	1980

2. (bases 1 to 800)  
Fujiyama, A., Hattori, M., Toyoda, A.,  
Tofukami, Y., Watanabe, H. and Sakaki, Y.  
Taylor, T. D., Yada, T.,

**TITLE**  
**JOURNAL**

Totoki, Y., Watanabe, H. and Sakaki, Y.  
Direct Submission  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suenhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: [champs@gsc.riken.go.jp](mailto:champs@gsc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT
<p>Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the BAC screening and was library hit #3909 of</p>

COMMENT  
Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

clone tracking errors.  
PRIMERS  
Sequencing: M13Rev  
TTPADP

```

VECTOR      : pKS145
R.Site 1   : Saci
R.Site 2   : Saci
Location/Qualifiers
source     1. .800

```

```
source
1. 000
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-081L22.R"
/sex="male"
```

```

BASE COUNT
ORIGIN
173 a 214 c 148 g 254 t 11 others
/clone_lib="PTB Chimpanzee Male BAC Library"
/cell_type="lymphoblast"
/sex="male"
/clone="PTB-081L2.R"
/taxon="Homo. 3236"

```

Query Match	1.0%;	Score 21;	DB 17;	Length 800;
Best Local Similarity	100.0%;			
Pred. No. 22;				

```

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 553 GAATACCTCCTCATCGTCTAC 573
      |||||
Db 575 GAATACCTCCTCATCGTCTAC 595

RESULT 4
LOCUS BM553311/c
DEFINITION AGENCOURT_6542482 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742946
5', mRNA sequence.
ACCESSION BM553311
VERSION BM553311.1 GI:18791932
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1061)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12761 row: o column: 11
High quality sequence stop: 682.
FEATURES
    .source
    1..1061
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:5742946"
        /clone_lib="NIH_MGC_119"
        /tissue_type="medulla"
        /lab_host="DH10B"
        /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
        Site_2: EcoRV (destroyed); RNA source normal medulla from
        anonymous male age 27. Library is oligo-dT primed and
        directionally cloned (EcoRV site is destroyed upon
        cloning). Average insert size 1.3 kb, insert size range
        0.9-3 kb. Library is normalized and enriched for
        full-length clones and was constructed by C. Gruber
        (Invitrogen). Research Genetics tracking code 013. Note:
        this is a NIH_MGC Library."
BASE COUNT 254 a 310 c 287 g 209 t
ORIGIN
Query Match 1.0%; Score 21; DB 13; Length 1061;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 TTAAGAACTATTCTTCGGA 33
      |||||
Db 897 TTAAGAACTATTCTTCGGA 877

RESULT 5
LOCUS AW200173
DEFINITION AW200173 Xenopus laevis oocyte Xenopus laevis cDNA clone
XENOPUS_SOURCE_ID:xlnc003015 5', similar to TR:035934 O35934 CLN3.
; mRNA sequence.
ACCESSION AW200173
VERSION AW200173.1 GI:6480742
KEYWORDS EST.
SOURCE African clawed frog.

```

```

ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 223)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
WashU Xenopus EST project, 1999
Unpublished (1999)
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: dalze08.xl
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: http://www.resgen.com/
Seq primer: -40RP from Gibco
High quality sequence stop: 125.
FEATURES
    .source
    1..223
        /organism="Xenopus laevis"
        /db_xref="taxon:8355"
        /clone="XENOPUS_SOURCE_ID:xlnc003015"
        /clone_lib="Xenopus laevis oocyte"
        /tissue_type="oocyte (stages 5 and 6)"
        /lab_host="Top-10 F"
        /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
        XhoI; cDNA was prepared from 2ug of poly A+ RNA.
        EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
        (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
        end. SS-library phagemids were prepared by mass excision
        from the original library and normalized by hybridization
        to biotinylated driver (prepared from the same library by
        PCR) to Cot-omega of 11. After removal of hybrids and
        excess driver by streptavidin sepharose chromatography,
        the ss-phagemids were made double stranded and
        electroporated into Top-10 F'. Original library
        constructed by Bruce Blumberg (Blumberg et al., 1991
        Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
        2923-2935). Normalized by Jihwan Song (Song, Cho and
        Blumberg, unpublished). Note: This is a Xenopus Gene
        Collection (XGC) library."
BASE COUNT 39 a 60 c 56 g 68 t
ORIGIN
Query Match 0.9%; Score 20; DB 10; Length 223;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 CAATCAGGAGAACTCGGATC 121
      |||||
Db 159 CAATCAGGAGAACTCGGATC 178

RESULT 6
LOCUS AW881153
DEFINITION QV0-OT0033-170400-197-c10 OT0033 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW881153
VERSION AW881153.1 GI:8043163
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



```

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 257)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV0-OT0033-170
400-197-cl08t3-2000-04-17&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 257.
FEATURES
    source
        Location/Qualifiers
            1..257
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="OT0033"
                /dev_stage="Adult"
                /note="Organ: ovary; Vector: puc18; Site 1: SmaI; Site 2:
                SmaI; A mini-library was made by cloning products derived
                from ORESTES PCR (U.S. Letters Patent application No. 196
                716 - Ludwig Institute for Cancer Research) profiles
                into the pUC 18 vector. Reverse transcription of tissue
                mRNA and cDNA amplification were performed under low
                stringency conditions."
BASE COUNT      86 a   30 c   57 g   84 t
ORIGIN
Query Match      0.9%; Score 20; DB 10; Length 257;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 956 GTATCAATATGCCAGCAAG 975
|||||
Db 83 GTATCAATATGCCAGCAAG 102

RESULT 7
BG158262/c
LOCUS      BG158262      260 bp      mRNA      linear      EST 08-MAR-2001
DEFINITION EM1_9_F06.b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION  BG158262
VERSION     BG158262.1 GI:12691912
KEYWORDS   EST.
SOURCE      sorghum.
    ORGANISM
        Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 260)
Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 257)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV0-OT0033-170
400-197-cl08t3-2000-04-17&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 257.
FEATURES
    source
        Location/Qualifiers
            1..257
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="OT0033"
                /dev_stage="Adult"
                /note="Organ: ovary; Vector: puc18; Site 1: SmaI; Site 2:
                SmaI; A mini-library was made by cloning products derived
                from ORESTES PCR (U.S. Letters Patent application No. 196
                716 - Ludwig Institute for Cancer Research) profiles
                into the pUC 18 vector. Reverse transcription of tissue
                mRNA and cDNA amplification were performed under low
                stringency conditions."
BASE COUNT      86 a   30 c   57 g   84 t
ORIGIN
Query Match      0.9%; Score 20; DB 10; Length 257;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 956 GTATCAATATGCCAGCAAG 975
|||||
Db 83 GTATCAATATGCCAGCAAG 102

RESULT 7
BG158262/c
LOCUS      BG158262      260 bp      mRNA      linear      EST 08-MAR-2001
DEFINITION EM1_9_F06.b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION  BG158262
VERSION     BG158262.1 GI:12691912
KEYWORDS   EST.
SOURCE      sorghum.
    ORGANISM
        Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 260)
Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.
An EST database from Sorghum: developing embryos
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860

```

```

Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence start: 5
High quality sequence stop: 116
POLYA-No.      Location/Qualifiers
    1..260
        /organism="Sorghum bicolor"
        /db_xref="taxon:4558"
        /clone_lib="Embryo 1 (EM1)"
        /note="Organ: Embryos germinated for 24 hr; Vector:
        pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2:
        EcoRI; The library was made from poly-A RNA in the cloning
        vector lambda Zap II. Clones to be sequenced were
        prepared by mass excision."
BASE COUNT      71 a   69 c   110 g   10 t
ORIGIN
Query Match      0.9%; Score 20; DB 12; Length 260;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1020 CCGCGGTCTCGCTGCTCTTG 1039
|||||
Db 160 CCGCGGTCTCGCTGCTCTTG 141

RESULT 8
AW199497/c
LOCUS      AW199497      410 bp      mRNA      linear      EST 27-JUN-2001
DEFINITION GA12608.x1 Xenopus laevis oocyte Xenopus laevis cDNA clone
XENOPUS_SOURCE_ID:xlnc003015 3', similar to SW:CLN3-CANPA Q29611
CLN3 PROTEIN. ;, mRNA sequence.
ACCESSION  AW199497
VERSION     AW199497.1 GI:6480066
KEYWORDS   EST.
SOURCE      African clawed frog.
    ORGANISM
        Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
            Xenopodinae; Xenopus.
1 (bases 1 to 410)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person,
B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: http://www.resgen.com/
Seq primer: -400P from Gibco
High quality sequence stop: 286.
FEATURES
    source
        Location/Qualifiers
            1..410
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"
                /clone_lib="XENOPUS_SOURCE_ID:xlnc003015"
                /clone_lib="Xenopus laevis oocyte"

```

```

/tissue_type="oocyte (stages 5 and 6)"
/lab_host="top-10 f"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI;
Library made from 2ug of poly A+ RNA.
EcoRI-XhoI cut cDNA was then ligated into Unizap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'
end. SS-library phagemids were prepared by mass excision
from the original library and normalized by hybridization
to biotinylated driver (prepared from the same library by
PCR) to Cot-omega of 11. After removal of hybrids and
excess driver by streptavidin sepharose chromatography,
the ss-phagemids were made double stranded and
electroporated into top-10 f. Original library
contruction by Bruce Blumberg (Blumberg et al., 1991
Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
2923-2935). Normalized by Jihwan Song (Song, Cho and
Blumberg, unpublished). Note: This is a Xenopus Gene
Collection (XGC) library."
BASE COUNT      121 a 101 c 112 g 76 t
ORIGIN
Query Match      0.9%; Score 20; DB 10; Length 410;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 CAATCAGGAACTGGATC 121
|||||
Db 373 CAATCAGGAACTGGATC 354

RESULT 9
LOCUS BE809917 419 bp mRNA linear EST 25-APR-2001
DEFINITION 217225 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE809917
VERSION BE809917.1 GI:10241029
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 419)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt.trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 75 row: G column: 10
Seq primer: ATTTAGTGACACTATAG.
Location/Qualifiers
1. 419
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"

/notes="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT      76 a 147 c 118 g 78 t
ORIGIN
Query Match      0.9%; Score 20; DB 12; Length 419;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 TCCTCTCGGAGCAGCCCTGC 43
|||||
Db 95 TCCTCTCGGAGCAGCCCTGC 114

RESULT 10
LOCUS BQ412487 463 bp mRNA linear EST 22-MAY-2002
DEFINITION GA_Ed0058F03r Gossypium arboreum 7-10 dpa fiber library Gossypium
arboreum cDNA clone GA_Ed0058F03r, mRNA sequence.
ACCESSION BQ412487
VERSION BQ412487.1 GI:21100174
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 463)
Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Total High Quality bases = 447
Seq primer: TAATACGACTCACTATAGGG
High quality sequence start: 2
High quality sequence stop: 464.
Location/Qualifiers
1. 463
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ed0058F03r"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="E. coli"
/notes="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT      114 a 144 c 65 g 140 t
ORIGIN
Query Match      0.9%; Score 20; DB 14; Length 463;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1489 TTCCATGCCATGCTCAAGTC 1508
|||||
Db 87 TTCCATGCCATGCTCAAGTC 106

RESULT 11
LOCUS BQ403470/c 472 bp mRNA linear EST 22-MAY-2002

```

```

DEFINITION  GA_Ed0058F03f Gossypium arboreum 7-10 dpa fiber library Gossypium
arboresum cdna clone GA_Ed0058F03f, mRNA sequence.
ACCESSION   B0403470
VERSION     B0403470.1  GI:21091157
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
ORGANISM    Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1. (bases 1 to 472)
AUTHORS     Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE       An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
JOURNAL     Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total Quality bases = 216
            Seq primer: TAATAGCACTCACTATAGGG
            High quality sequence start: 3
            High quality sequence stop: 446.
FEATURES    Location/Qualifiers
            source
            1..472
            /organism="Gossypium arboreum"
            /strain="AKA"
            /cultivar="8400"
            /db_xref="taxon:29729"
            /clone="GA_Ed0058F03f"
            /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
            /tissue_type="Fibers isolated from bolls harvested 7-10
            dpa"
            /lab_host="E. coli"
            /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT  137 a 80 c 131 g 124 t
ORIGIN
Query Match      0.9%; Score 20; DB 14; Length 472;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1489  TTCCATGCCATGCTCAAGTC 1508
            |||||
Db  378  TTCCATGCCATGCTCAAGTC 359

RESULT 12
BF654963
LOCUS       BF654963                473 bp mRNA linear EST 25-APR-2001
DEFINITION  279329 MARC 3BOV Bos taurus cdna 5', mRNA sequence.
ACCESSION   BF654963
VERSION     BF654963.1  GI:11920095
KEYWORDS    EST.
SOURCE      cow.
ORGANISM    Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1. (bases 1 to 473)
AUTHORS     Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Partea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
TITLE       Sequence evaluation of four pooled-tissue normalized bovine cdna
            libraries and construction of a gene index for cattle
JOURNAL     Genome Res. 11 (4), 626-630 (2001)
MEDLINE     21180013

```

```

COMMENT      Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@mail.marc.usda.gov
            Single pass sequencing. Bases called and alt.trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAACAGCTATGACCAT
            BACKWARD: GTTTCCTCCAGTCACGACG
            Plate: 74 row: G column: 8
            Seq primer: ATTTAGTGACACTATAG.
FEATURES    Location/Qualifiers
            source
            1..473
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="MARC 3BOV"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
            Library made from pooled tissue from marrow, alveolar
            macrophage, ovary, fetal semitendinosus muscle, and fetal
            longissimus muscle."
BASE COUNT  82 a 175 c 125 g 91 t
ORIGIN
Query Match      0.9%; Score 20; DB 12; Length 473;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  24  TCTTCTCGGACGACCCCTGC 43
            |||||
Db  383  TCTTCTCGGACGACCCCTGC 402

RESULT 13
AI718392/c
LOCUS       AI718392                509 bp mRNA linear EST 10-JUN-1999
DEFINITION  as5906.x1 Barstead colon HPLR87 Homo sapiens cdna clone
            IMAGE:2333050 3', similar to SW:Y281_HUMAN Q92556 HYPOTHETICAL
            PROTEIN KIAA0281.J, mRNA sequence.
ACCESSION   AI718392
VERSION     AI718392.1  GI:5035648
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1. (bases 1 to 509)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le.N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-NCI human EST Project
JOURNAL     Unpublished (1997)
COMMENT     Contact: Willson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -400P from Gibco
            High quality sequence stop: 220.
FEATURES    Location/Qualifiers
            source
            1..509
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2333050"

```

```

/clone_lib="Barstead colon HPLRB7"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/notes="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I oligo (dT) primer [5'
TGTTAGCAATCGAAGTGGAGCGCCCTTTTGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[5' AATTCACGTAAT 3' and 5' ATTACTAGTG 3'], digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT73 vector. Library constructed by Bob
Barstead."
BASE COUNT      114 a 123 c 140 g 132 t
ORIGIN

Query Match      0.9%; Score 20; DB 9; Length 509;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2089 GACAAATGGGTGAGTGCCTC 2108
      |||||
Db 369 GACAAATGGGTGAGTGCCTC 350

RESULT 14
BE809904
LOCUS      217039 MARC 2BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
DEFINITION
ACCESSION  BE809904
VERSION     BE809904.1 GI:10241016
KEYWORDS   EST.
SOURCE     cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 519)
AUTHORS    Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
            Keefe,J.W.
TITLE      Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL
MEDLINE
COMMENT    Contact: Smith.TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -mismatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 75 row: D column: 10
Seq primer: ATTTAGTGACACTATAG.
FEATURES
            source
            Location/Qualifiers
            1..519
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="MARC 2BOV"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
            Library made from pooled tissue from testis, thymus,
            semitendinosus muscle, longissimus muscle, pancreas,
            adrenal, and endometrium."
BASE COUNT      85 a 181 c 146 g 107 t

/clone_lib="Barstead colon HPLRB7"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/notes="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st
strand cDNA was primed with a Not I oligo (dT) primer [5'
TGTTAGCAATCGAAGTGGAGCGCCCTTTTGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[5' AATTCACGTAAT 3' and 5' ATTACTAGTG 3'], digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT73 vector. Library constructed by Bob
Barstead."
BASE COUNT      114 a 123 c 140 g 132 t
ORIGIN

Query Match      0.9%; Score 20; DB 9; Length 509;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2089 GACAAATGGGTGAGTGCCTC 2108
      |||||
Db 369 GACAAATGGGTGAGTGCCTC 350

RESULT 14
BE809904
LOCUS      217039 MARC 2BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
DEFINITION
ACCESSION  BE809904
VERSION     BE809904.1 GI:10241016
KEYWORDS   EST.
SOURCE     cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 519)
AUTHORS    Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
            Keefe,J.W.
TITLE      Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL
MEDLINE
COMMENT    Contact: Smith.TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -mismatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 75 row: D column: 10
Seq primer: ATTTAGTGACACTATAG.
FEATURES
            source
            Location/Qualifiers
            1..519
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="MARC 2BOV"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
            Library made from pooled tissue from testis, thymus,
            semitendinosus muscle, longissimus muscle, pancreas,
            adrenal, and endometrium."
BASE COUNT      85 a 181 c 146 g 107 t

Query Match      0.9%; Score 20; DB 12; Length 519;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 TCCTCTCGGAGCAGCCCTGC 43
      |||||
Db 95 TCCTCTCGGAGCAGCCCTGC 114

RESULT 15
BI847597
LOCUS      469747 MARC 2BOV Bos taurus cDNA 5', mRNA linear EST 04-OCT-2001
DEFINITION
ACCESSION  BI847597
VERSION     BI847597.1 GI:15960116
KEYWORDS   EST.
SOURCE     cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 523)
AUTHORS    Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and
            Keefe,J.W.
TITLE      Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
JOURNAL
MEDLINE
COMMENT    Contact: Smith.TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -mismatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 94 row: D column: 18
Seq primer: ATTTAGTGACACTATAG.
FEATURES
            source
            Location/Qualifiers
            1..523
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="MARC 2BOV"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
            Library made from pooled tissue from testis, thymus,
            semitendinosus muscle, longissimus muscle, pancreas,
            adrenal, and endometrium."
BASE COUNT      86 a 180 c 150 g 107 t
ORIGIN

Query Match      0.9%; Score 20; DB 13; Length 523;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 TCCTCTCGGAGCAGCCCTGC 43
      |||||
Db 96 TCCTCTCGGAGCAGCCCTGC 115

Search completed: May 23, 2003, 08:42:26
Job time : 2065 secs

```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 13:03:17 ; Search time 30 seconds  
(without alignments)  
4195.709 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 712

Sequence: 1 atgcaaatgaattataaaag.....aagagctgaattgatctaa 2139

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool/US1008355/runat\_16052003\_110427\_9763/app.query.fasta.1.2311  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=olin2p.ra -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US1008355\_@CGN\_1\_1\_33\_@runat\_16052003\_110427\_9763 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	1.4	213	3	US-08-523-373-22
2	10	1.4	214	3	US-08-523-373-23
3	10	1.4	215	3	US-08-523-373-24
4	10	1.4	344	1	US-08-657-192-3
5	10	1.4	344	3	US-08-523-373-5
6	10	1.4	392	3	US-08-523-373-6
7	10	1.4	532	3	US-08-657-192-9
8	10	1.4	532	3	US-08-523-373-7
9	10	1.4	537	1	US-08-657-192-15
10	9	1.3	15	2	US-08-332-562A-67
11	9	1.3	311	4	US-09-134-001C-3729
12	9	1.3	984	2	US-08-673-789-9

13	9	1.3	984	2	US-08-449-645A-19	Sequence 19, Appl
14	9	1.3	984	2	US-08-702-367A-19	Sequence 19, Appl
15	9	1.3	984	5	PCT-US95-04681-19	Sequence 19, Appl
16	8	1.1	116	1	US-08-434-705B-8	Sequence 8, Appl
17	8	1.1	116	2	US-09-086-201-8	Sequence 8, Appl
18	8	1.1	396	4	US-08-860-656B-2	Sequence 2, Appl
19	8	1.1	396	4	US-08-860-656B-3	Sequence 3, Appl
20	8	1.1	637	2	US-08-426-125-10	Sequence 10, Appl
21	8	1.1	637	2	US-08-455-355-10	Sequence 10, Appl
22	8	1.1	1170	4	US-09-749-588-2	Sequence 2, Appl
23	8	1.1	1209	4	US-09-749-588-4	Sequence 4, Appl
24	7	1.0	13	4	US-08-602-999A-89	Sequence 89, Appl
25	7	1.0	13	4	US-08-278-865-89	Sequence 89, Appl
26	7	1.0	13	4	US-09-500-124-89	Sequence 10, Appl
27	7	1.0	20	4	US-08-860-656B-10	Sequence 10, Appl
28	7	1.0	31	4	US-08-602-999A-57	Sequence 57, Appl
29	7	1.0	31	4	US-08-278-865-57	Sequence 57, Appl
30	7	1.0	31	4	US-09-500-124-57	Sequence 57, Appl
31	7	1.0	39	4	US-09-172-841-57	Sequence 57, Appl
32	7	1.0	43	3	US-08-899-578-5	Sequence 38, Appl
33	7	1.0	45	2	US-08-766-858A-38	Sequence 21, Appl
34	7	1.0	51	4	US-09-150-460B-21	Sequence 21, Appl
35	7	1.0	53	4	US-08-905-223-496	Sequence 496, App
36	7	1.0	98	1	US-08-308-086-4	Sequence 4, Appl
37	7	1.0	98	2	US-08-479-078-5	Sequence 5, Appl
38	7	1.0	98	4	US-08-975-040-22	Sequence 22, Appl
39	7	1.0	99	1	US-08-202-389-38	Sequence 38, Appl
40	7	1.0	101	2	US-08-574-959A-5	Sequence 5, Appl
41	7	1.0	101	4	US-09-357-014-5	Sequence 5, Appl
42	7	1.0	102	2	US-08-820-754-24	Sequence 24, Appl
43	7	1.0	102	3	US-08-956-652-24	Sequence 24, Appl
44	7	1.0	102	3	US-08-956-869-24	Sequence 24, Appl
45	7	1.0	102	3	US-08-948-547-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1

US-08-523-373-22  
; Sequence 22, Application US/08523373  
; Patent No. 6037145  
; GENERAL INFORMATION:  
; APPLICANT: Yabuta, Masayuki  
; APPLICANT: Ohsuye, Kazuhiro  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/523,373  
; FILING DATE: 05-SEP-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-238595  
; FILING DATE: 07-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-296028  
; FILING DATE: 07-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 001560-251  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 213 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-22

Alignment Scores:  
Pred. No.: 0.351 Length: 213  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 3 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-523-373-22 (1-213)  
QY 1930 ACGGCGGTAACTCCGGTAGCCCGTATTC 1959  
Db 165 ThrGlyGlyAsnSerGlySerProValPhe 174

RESULT 2  
US-08-523-373-23  
Sequence 23 Application US/08523373  
Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohsuye, Kazuhiro  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 214 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-23

Alignment Scores:  
Pred. No.: 0.351 Length: 213  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 3 Gaps: 0

Pred. No.: 0.351 Length: 214  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 3 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-523-373-23 (1-214)  
QY 1930 ACGGCGGTAACTCCGGTAGCCCGTATTC 1959  
Db 165 ThrGlyGlyAsnSerGlySerProValPhe 174

RESULT 3  
US-08-523-373-24  
Sequence 24 Application US/08523373  
Patent No. 6037145  
GENERAL INFORMATION:  
APPLICANT: Yabuta, Masayuki  
APPLICANT: Ohsuye, Kazuhiro  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 215 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-24

Alignment Scores:  
Pred. No.: 0.351 Length: 215  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 3 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-523-373-24 (1-215)  
QY 1930 ACGGCGGTAACTCCGGTAGCCCGTATTC 1959  
Db 165 ThrGlyGlyAsnSerGlySerProValPhe 174



RESULT 4  
US-08-657-192-3  
; Sequence 3, Application US/08657192  
; Patent No. 5747321  
; GENERAL INFORMATION:  
; APPLICANT: YABUTA, Masayuki  
; APPLICANT: OHSUYE, Kazuhiro  
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22131-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 03-JUN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION NUMBER: JP 7-170086  
; FILING DATE: 02-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meuth, Donna M.  
; REGISTRATION NUMBER: 36,607  
; REFERENCE/DOCKET NUMBER: 001560-264  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-5620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-657-192-3

Alignment Scores:  
Pred. No.: 0.332 Length: 344  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-657-192-3 (1-344)  
QY 1930 ACGGGCGTAACCGGTAGCCCGTATTC 1959  
|||||  
Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 5  
US-08-523-373-5  
; Sequence 5, Application US/08523373  
; Patent No. 6037145  
; GENERAL INFORMATION:  
; APPLICANT: Yabuta, Masayuki  
; APPLICANT: Ohsuye, Kazuhiro  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 05-SEP-1995  
; CLASSIFICATION: 435

STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/523,373  
FILING DATE: 05-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-238595  
FILING DATE: 07-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-296028  
FILING DATE: 07-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meuth, Donna M.  
REGISTRATION NUMBER: 36,607  
REFERENCE/DOCKET NUMBER: 001560-251  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-523-373-5

Alignment Scores:  
Pred. No.: 0.332 Length: 344  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 3 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-523-373-5 (1-344)  
QY 1930 ACGGGCGTAACCGGTAGCCCGTATTC 1959  
|||||  
Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 6  
US-08-523-373-6  
; Sequence 6, Application US/08523373  
; Patent No. 6037145  
; GENERAL INFORMATION:  
; APPLICANT: Yabuta, Masayuki  
; APPLICANT: Ohsuye, Kazuhiro  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22314-3187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 05-SEP-1995  
; CLASSIFICATION: 435

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIOR APPLICATION DATA: JP 6-296028
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 392 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373-6

Alignment Scores:
Pred. No.: 0.327 Length: 392
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 3 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-523-373-6 (1-392)
QY 1930 ACGGCGGTAACTCCGGTAGCCCGTATTC 1959
Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 7
US-08-657-192-9
; Sequence 9, Application US/08657192
; Patent No. 5747321
; GENERAL INFORMATION:
; APPLICANT: YABUTA, Masayuki
; APPLICANT: OHSUYE, Kazuhiro
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; TITLE OF INVENTION: PROTEASES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/657,192
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-170086
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-657-192-9

Alignment Scores:
Pred. No.: 0.315 Length: 532
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Query Match: 1.40% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-657-192-9 (1-532)
QY 1930 ACGGCGGTAACTCCGGTAGCCCGTATTC 1959
Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 8
US-08-523-373-7
; Sequence 7, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohsuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373-7

Alignment Scores:
Pred. No.: 0.315 Length: 532
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
```

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 3 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-523-373-7 (1-532)

QY 1930 ACGGGCGGTACTCGGTAGCCCGTATTC 1959  
 |||||  
 Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 9  
 US-08-657-192-15  
 ; Sequence 15, Application US/08657192  
 ; Patent No. 5747321  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YABUTA, Masayuki  
 ; APPLICANT: OHSUYE, Kazuhiro  
 ; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8  
 ; TITLE OF INVENTION: PROTEASES  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
 ; STREET: P.O. Box 1404  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: United States  
 ; ZIP: 22313-1404  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION NUMBER: US/08/657,192  
 ; FILING DATE: 03-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 7-170086  
 ; FILING DATE: 02-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meuth, Donna M.  
 ; REGISTRATION NUMBER: 36,607  
 ; REFERENCE/DOCKET NUMBER: 001560-264  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 836-6620  
 ; TELEFAX: (703) 836-2021  
 ; INFORMATION FOR SEQ ID NO: 15:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 537 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-657-192-15

Alignment Scores:  
 Pred. No.: 0.314 Length: 537  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-657-192-15 (1-537)

QY 1930 ACGGGCGGTACTCGGTAGCCCGTATTC 1959  
 |||||  
 Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 10  
 US-08-332-562A-67  
 ; Sequence 67, Application US/08332562A  
 ; Patent No. 5985599

GENERAL INFORMATION:  
 ; APPLICANT: MCKENZIE, Ian F.C.  
 ; APPLICANT: HOGARTH, Mark P.  
 ; APPLICANT: HIBBS, Margaret L.  
 ; APPLICANT: SCOTT, Bernadette M.  
 ; APPLICANT: BONADONNA, Lisa  
 ; APPLICANT: HULETT, Mark D.  
 ; TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN  
 ; NUMBER OF SEQUENCES: 136  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W., Suite 500  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20007-5109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION NUMBER: US/08/332,562A  
 ; FILING DATE: 31-OCT-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/896,457  
 ; FILING DATE: 27-MAY-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 54270/119/GRHA  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202)672-5300  
 ; TELEFAX: (202)672-5399  
 ; TELEX: 904136  
 ; INFORMATION FOR SEQ ID NO: 67:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 15 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-332-562A-67

Alignment Scores:  
 Pred. No.: 4.44 Length: 15  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.26% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x US-08-332-562A-67 (1-15)

QY 1930 ACGGGCGGTACTCGGTAGCCCGTA 1956  
 |||||  
 Db 5 ThrGlyGlyAsnSerGlySerProVal 13

RESULT 11  
 US-09-134-001C-3729  
 ; Sequence 3729, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14



;; CITY: Thousand Oaks  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 91320  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/702,367A  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Winter, Robert B.  
;; REFERENCE/DOCKET NUMBER: A-287  
;; INFORMATION FOR SEQ ID NO: 19:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 984 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-702-367A-19  
  
Alignment Scores:  
Pred. No.: 2.68 Length: 984  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.26% Indels: 0  
DB: 2 Gaps: 0  
  
US-10-008-355-1 (1-2139) x US-08-702-367A-19 (1-984)  
  
Qy 25 CTTCTCGGAGCAGCCCTGCTGTGGGT 51  
Db 556 LeuLeuGlyAlaAlaLeuLeuGly 564  
  
RESULT 15  
PCT-US95-04681-19  
; Sequence 19: Application PC/TUS9504681  
; GENERAL INFORMATION:  
; APPLICANT: Fox, Gary M.  
; TITLE OF INVENTION: EPH-Like Receptor Protein Tyrosine  
; TITLE OF INVENTION: Kinases  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Patent Operations/RBW  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/04681  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Winter, Robert B.  
; REFERENCE/DOCKET NUMBER: A-287  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 984 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

PCT-US95-04681-19  
  
Alignment Scores:  
Pred. No.: 2.68 Length: 984  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.26% Indels: 0  
DB: 5 Gaps: 0  
  
US-10-008-355-1 (1-2139) x PCT-US95-04681-19 (1-984)  
  
Qy 25 CTTCTCGGAGCAGCCCTGCTGTGGGT 51  
Db 556 LeuLeuGlyAlaAlaLeuLeuGly 564  
  
Search completed: May 16, 2003, 13:17:28  
Job time : 44 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:38:22 ; Search time 113.5 Seconds  
(without alignments)  
5022.433 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 712

Sequence: 1 atgcaaatgaattataaaag.....aagagctgaagtgtatctaa 2139

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1687582

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2.1/USPPO.spool/US10008355/runat\_16052003\_110425\_9716/app\_query.fasta\_1.2311  
-DB=A\_Geneseq\_101002 -QFMT=fastan -SUFFIX=olin2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human4.0.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355 -CGN\_1\_1\_153 -runat\_16052003\_110425\_9716 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -LONGLOS -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_101002:\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	712	100.0	712	23	AAO15205 Porphyromonas ging
2	52	7.3	52	23	AAO15206 Porphyromonas ging
3	21	2.9	26	23	AAO15221 Porphyromonas ging
4	10	1.4	10	23	AAO15228 Porphyromonas ging
5	10	1.4	52	23	AAO15207 Staphylococcus aur
6	10	1.4	213	17	AAR91042 V8 mature protease
7	10	1.4	214	17	AAR91043 V8 mature protease
8	10	1.4	215	17	AAR91044 V8 mature protease
9	10	1.4	336	13	AAR29644 Protease from S. A
10	10	1.4	344	17	AAR91033 Beta-galactosidase
11	10	1.4	344	18	AAW22218 Protein encoded by
12	10	1.4	357	13	AAR26842 Protease from S. A
13	10	1.4	392	17	AAR91034 Beta-galactosidase
14	10	1.4	532	17	AAR91035 Recombinant V8 pro
15	10	1.4	532	18	AAW22219 Protein encoded by
16	10	1.4	537	18	AAW22220 Porphyromonas ging
17	9	1.3	9	23	AAO15229 Propionibacterium
18	9	1.3	76	22	AAU63245 S. epidermidis ope
19	9	1.3	282	22	AAU61635 Staphylococcus epi
20	9	1.3	311	23	ABP38884 Omega-conopeptide
21	8	1.1	27	23	ABB96842 Human CART peptide
22	8	1.1	55	21	AA35780 Omega-conopeptide
23	8	1.1	73	23	AB96626 Human CART peptide
24	8	1.1	75	21	AA35779 Human CART peptide
25	8	1.1	80	22	AAU50283 Propionibacterium
26	8	1.1	88	21	AA35772 Human truncated CA
27	8	1.1	89	22	AAU66041 Propionibacterium
28	8	1.1	116	17	AAW06814 Human cocaine and
29	8	1.1	116	19	AAW78906 Human cocaine and
30	8	1.1	116	20	AA45180 Human cocaine and
31	8	1.1	116	20	AAW81341 Human polyptide
32	8	1.1	138	22	AAO07965 Human protein sequ
33	8	1.1	152	22	AA94914 Human digestive sy
34	8	1.1	159	22	AA92139 Novel human diago
35	8	1.1	173	22	ABG18197 Drosophila melanog
36	8	1.1	211	22	ABG67468 Arabidopsis transc
37	8	1.1	236	23	AAU93181 Novel human diago
38	8	1.1	306	22	ABG27777 Human T2R09 amino
39	8	1.1	312	22	AA87739 Streptococcus poly
40	8	1.1	335	23	ABP30343 Streptococcus poly
41	8	1.1	340	23	ABP28097 Streptococcus poly
42	8	1.1	340	23	ABP29857 Escherichia coli p
43	8	1.1	395	22	AB52640 Interleukin-2 rece
44	8	1.1	396	17	AAR97569 Human IL-2R-associ
45	8	1.1	413	22	ABB11836

ALIGNMENTS

RESULT 1  
AAO15205  
ID AAO15205 standard; Protein; 712 AA.  
XX  
AC AAO15205;  
XX  
DT 05-SEP-2002 (first entry)  
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).  
XX  
KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
KW periodontitis.  
XX  
OS Porphyromonas gingivalis.  
PN WO200238742-A2.  
XX  
PD 16-MAY-2002.





Db 501 LeuLysSerMetAspLysGluLysPheAlaLysAlaLysAlaLysAlaLysAspProAlaValGlu 520  
QY 1561 CTTTCCAGAGCGTAATAGCTGCTCGCGCTATTGAGGCCGATGCGATGGCCAAATGCC 1620  
Db 521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540  
QY 1621 TATGCCATTGAGAGGCGAGCGCTTTTCTTCCCGGTTTGGCTGAGATGTACCCCGGA 1680  
Db 541 TyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrProGly 560  
QY 1681 CGTCTCTGCGGACGATGCCAACTTCCACATGCTGATGAGCTACGCTCCATCAAGGA 1740  
Db 561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580  
QY 1741 TATGAACGCGAGGCGTGGCTGTGACAACTATCATACGAGCAAGGCGGATTTGGAG 1800  
Db 581 TyrGluProGlnAspGlyAlaTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600  
QY 1801 AAGCAGATCTTAAGACCGATGAGTTTGGCTACAGGAGATATCTTCGACCTCTTCGGC 1860  
Db 601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620  
QY 1861 ACCAAATATGCTGCTATGCGGAGAACGCTCAGCTCCATATCGCTTTCCTATCGAAC 1920  
Db 621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640  
QY 1921 AACGACATCACGGCGGTAACTCCGGTAGCCCGCTATTTCGATAAGACGGCGCTGATC 1980  
Db 641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660  
QY 1981 GGTCTTCTGATGGCACTGGAGCTATGATGAGTGGTGACATCGAGTTCGAACCCGAT 2040  
Db 661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680  
QY 2041 CTGACGCGCAATCAGCGTGACATCGCTACGCTTCTTTCATGATTGACAAATGGGT 2100  
Db 681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700  
QY 2101 CAGTCCCCCTCTCATCCAGAGCTGAAGTTGATC 2136  
Db 701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712  
RESULT 2  
AAOI5206  
ID AAO15206 standard; Protein: 52 AA.  
XX AC AAO15206;  
XX AC AAO15206;  
DT 05-SEP-2002 (first entry)  
XX DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.  
XX DE Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
XX KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
XX KW periodontitis.  
XX OS Porphyromonas gingivalis.  
XX OS WO200238742-A2.  
XX PN 16-MAY-2002.  
XX PD 08-NOV-2001; 2001WO-US46782.  
XX PF 08-NOV-2000; 2000US-246827P.  
XX PR (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX PA Travis J, Potempa JS, Banbula A, Bugno M;  
XX PI WPI; 2002-490075/52.  
XX DR Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
PT

PT the dipeptidylpeptidase for protecting an animal from periodontal  
PT disease caused by Porphyromonas gingivalis -  
XX Example 6; Fig 5; 65pp; English.  
PS  
XX The invention comprises the amino acid and coding sequence of the  
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
CC second and third amino acids from the N-terminal end of a target peptide.  
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a  
CC substituent on the alpha-carbon atom of the second amino acid from the  
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
CC reducing the growth of a bacterium and protecting an animal from a  
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
CC or periodontitis). The present amino acid sequence represents the  
CC C-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7  
CC (DPP-7) enzyme.  
XX SQ Sequence 52 AA;  
XX Alignment Scores:  
Pred. No.: 1.17e-44 Length: 52  
Score: 52.00 Matches: 52  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.30% Indels: 0  
DB: 23 Gaps: 0  
US-10-008-355-1 (1-2139) x AAO15206 (1-52)  
QY 1930 ACGGCGGTAACTCCGCTAGCCCGCTATTCGATAAGAACGCCGCTCTGATCGTCTTGT 1989  
Db 1 ThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla 20  
QY 1990 TTCCATGCGCAACTCGGAGGAGCTATGAGTGGTGACATCGAGTTCGAACCGGATCTCGAGCGC 2049  
Db 21 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 40  
QY 2050 ACAATCAGCTGGAGATCCGCTAGCTTCTTCTCATG 2085  
Db 41 ThrIleSerValAspIleArgTyrValLeuPheMet 52  
RESULT 3  
AAOI5221  
ID AAO15221 standard; Peptide: 26 AA.  
XX AC AAO15221;  
XX AC AAO15221;  
DT 05-SEP-2002 (first entry)  
XX DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) N-terminal region.  
XX DE Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
XX KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
XX KW periodontitis.  
XX OS Porphyromonas gingivalis.  
XX OS WO200238742-A2.  
XX PN 16-MAY-2002.  
XX PD 08-NOV-2001; 2001WO-US46782.  
XX PF 08-NOV-2000; 2000US-246827P.  
XX PR (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX PA Travis J, Potempa JS, Banbula A, Bugno M;  
XX PI WPI; 2002-490075/52.  
XX DR  
XX

PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
 PT the dipeptidylpeptidase for protecting an animal from periodontal  
 PT disease caused by Porphyromonas gingivalis  
 XX  
 PS Example 6; Page 30; 65pp; English.  
 XX  
 CC The invention comprises the amino acid and coding sequence of the  
 CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
 CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
 CC second and third amino acids from the N-terminal end of a target peptide.  
 CC The DPP-7 target peptide has an aliphatic or aromatic residue as a  
 CC substituent on the alpha-carbon atom of the second amino acid from the  
 CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
 CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
 CC reducing the growth of a bacterium and protecting an animal from a  
 CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
 CC or periodontitis). The present amino acid sequence represents the  
 CC N-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7  
 CC (DPP-7) enzyme.  
 XX  
 SQ Sequence 26 AA;  
 Alignment Scores:  
 Pred. No.: 2,33e-12 Length: 26  
 Score: 21.00 Matches: 21  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 2,95% Indels: 0  
 DB: 23 Gaps: 0  
 US-10-008-355-1 (1-2139) x AAO15221 (1-26)  
 QY 82 ATGTGGCTCCCTCAACGAACTCAATCAGGAGAACTCTGGATCGAATGCGTGAGCTCGGCTTT 141  
 DB 6 MettripleLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPhe 25  
 QY 142 ACG 144  
 DB 26 Thr 26  
 RESULT 4  
 AAO15228  
 ID AAO15228 standard; Peptide; 10 AA.  
 XX  
 AC AAO15228;  
 XX  
 DT 05-SEP-2002 (first entry)  
 XX  
 DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) active site 1.  
 XX  
 KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage; active site;  
 KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
 KW periodontitis.  
 XX  
 OS Porphyromonas gingivalis.  
 XX  
 WO200238742-A2.  
 XX  
 PN 16-MAY-2002.  
 XX  
 PF 08-NOV-2001; 2001WO-US46782.  
 XX  
 PR (UYGE-) UNIV GEORGIA RES FOUND INC.  
 XX  
 PA Travis J, Potempa JS, Banbula A, Bugno M;  
 PI WPI; 2002-490075/52.  
 XX  
 DR Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
 PT the dipeptidylpeptidase for protecting an animal from periodontal  
 PT disease caused by Porphyromonas gingivalis

XX Claim 5; Page 32; 65pp; English.  
 PS  
 XX The invention comprises the amino acid and coding sequence of the  
 CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
 CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
 CC second and third amino acids from the N-terminal end of a target peptide.  
 CC The DPP-7 target peptide has an aliphatic or aromatic residue as a  
 CC substituent on the alpha-carbon atom of the second amino acid from the  
 CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
 CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
 CC reducing the growth of a bacterium and protecting an animal from a  
 CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
 CC or periodontitis). The present amino acid sequence represents an active  
 CC site region of the Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7)  
 CC enzyme.  
 XX  
 SQ Sequence 10 AA;  
 Alignment Scores:  
 Pred. No.: 0.72 Length: 10  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 23 Gaps: 0  
 US-10-008-355-1 (1-2139) x AAO15228 (1-10)  
 QY 1930 ACGGCGGTAACTCCGGTAGCCCGGTATTC 1959  
 DB 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
 RESULT 5  
 AAO15207  
 ID AAO15207 standard; Protein; 52 AA.  
 XX  
 AC AAO15207;  
 XX  
 DT 05-SEP-2002 (first entry)  
 XX  
 DE Staphylococcus aureus V8 endopeptidase C-terminal region.  
 XX  
 KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;  
 KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
 KW periodontitis; V8 endopeptidase.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 WO200238742-A2.  
 XX  
 PN 16-MAY-2002.  
 XX  
 PF 08-NOV-2001; 2001WO-US46782.  
 XX  
 PR 08-NOV-2000; 2000US-246827P.  
 XX  
 PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
 XX  
 PI Travis J, Potempa JS, Banbula A, Bugno M;  
 XX  
 DR WPI; 2002-490075/52.  
 XX  
 PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
 PT the dipeptidylpeptidase for protecting an animal from periodontal  
 PT disease caused by Porphyromonas gingivalis  
 XX  
 PS Example 6; Fig 5; 65pp; English.  
 XX  
 CC The invention comprises the amino acid and coding sequence of the  
 CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
 CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
 CC second and third amino acids from the N-terminal end of a target peptide.

CC The DPP-7 target peptide has an aliphatic or aromatic residue as a  
 CC substituent on the alpha-carbon atom of the second amino acid from the  
 CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
 CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
 CC reducing the growth of a bacterium and protecting an animal from a  
 CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
 CC or periodontitis). The present amino acid sequence represents the  
 CC C-terminal region of the Staphylococcus aureus V8 endopeptidase enzyme.

XX Sequence 52 AA;

Alignment Scores:  
 Pred. No.: 0.618 Length: 52  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 23 Gaps: 0

US-10-008-355-1 (1-2139) x AAO15207 (1-52)

QY 1930 ACGGCGGTAACTCCGTCGTCAGCCCGTATTC 1959

Db 1 ThrGlyGlyAsnSerGlySerProValPhe 10

RESULT 6

AAR91042

ID AAR91042 standard; Peptide; 213 AA.

XX AAR91042;

XX 23-MAY-1996 (first entry)

XX V8 mature protease (aal-213).

XX Linker peptide; V8 protease; Staphylococcus aureus;

XX recombinant protein; fusion protein; beta-galactosidase;

XX Escherichia coli; transposon Tn903;

XX aminoglycoside 3'-phosphotransferase.

XX Staphylococcus aureus strain V8 (ATCC 27733).

XX EP700995-A2.

XX 13-MAR-1996.

XX 06-SEP-1995; 95EP-0306235.

XX 07-NOV-1994; 94JP-0296028.

XX 07-SEP-1994; 94JP-0238595.

XX (SUNR ) SUNTORY LTD.

XX Ohsuye K, Yabuta M;

XX WPI; 1996-141021/15.

XX Prodn. of recombinant polypeptide(s) - using host cells transformed

XX with a gene coding for the desired polypeptide fused to a protective

XX polypeptide

XX Example 8; Fig 14a; 44pp; English.

XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A

XX and C are protective polypeptides, B is Staphylococcus aureus mature

XX V8 protease lacking the C-terminal repeat region, and L is a linker

XX (AAR91032). It was produced as an inclusion body in Escherichia coli

XX host cells, and was cleaved using the E. coli Ompr protease to

XX yield active V8 protease. Extension of the C-terminal end of the

XX V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)

XX i.e. up to residue Phe-215, also resulted in the formation of

XX inclusion bodies in E. coli. Any further extension gave a soluble

XX product which exhibited protease activity that repressed growth of

XX CC

CC the host cells.

XX Sequence 213 AA;

SQ Alignment Scores:

Pred. No.: 0.542 Length: 213  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 17 Gaps: 0

US-10-008-355-1 (1-2139) x AAR91042 (1-213)

QY 1930 ACGGCGGTAACTCCGTCGTCAGCCCGTATTC 1959

Db 165 ThrGlyGlyAsnSerGlySerProValPhe 174

RESULT 7

AAR91043

ID AAR91043 standard; Peptide; 214 AA.

XX AAR91043;

XX 23-MAY-1996 (first entry)

XX V8 mature protease (aal-214).

XX Linker peptide; V8 protease; Staphylococcus aureus;

XX recombinant protein; fusion protein; beta-galactosidase;

XX Escherichia coli; transposon Tn903;

XX aminoglycoside 3'-phosphotransferase.

XX Staphylococcus aureus strain V8 (ATCC 27733).

XX EP700995-A2.

XX 13-MAR-1996.

XX 06-SEP-1995; 95EP-0306235.

XX 07-NOV-1994; 94JP-0296028.

XX 07-SEP-1994; 94JP-0238595.

XX (SUNR ) SUNTORY LTD.

XX Ohsuye K, Yabuta M;

XX WPI; 1996-141021/15.

XX Prodn. of recombinant polypeptide(s) - using host cells transformed

XX with a gene coding for the desired polypeptide fused to a protective

XX polypeptide

XX Example 8; Fig 14b; 44pp; English.

XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A

XX and C are protective polypeptides, B is Staphylococcus aureus mature

XX V8 protease lacking the C-terminal repeat region, and L is a linker

XX (AAR91032). It was produced as an inclusion body in Escherichia coli

XX host cells, and was cleaved using the E. coli Ompr protease to

XX yield active V8 protease. Extension of the C-terminal end of the

XX V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)

XX i.e. up to residue Phe-215, also resulted in the formation of

XX inclusion bodies in E. coli. Any further extension gave a soluble

XX product which exhibited protease activity that repressed growth of

XX the host cells.

XX SQ Sequence 214 AA;

Alignment Scores:

Pred. No.: 0.541 Length: 214

Score: 10.00 Matches: 10

Percent Similarity: 100.00%    Conservative: 0  
 Best Local Similarity: 100.00%    Mismatches: 0  
 Query Match: 1.40%    Indels: 0  
 DB: 17    Gaps: 0

US-10-008-355-1 (1-2139) x AAR91043 (1-214)

QY 1930 ACGGCGGTAACTCCGGTAGCCCGGTATTC 1959  
 |||||  
 Db 165 ThrGlyGlyAsnSerGlySerProValphe 174

## RESULT 8

AAR91044  
 ID AAR91044 standard; Peptide: 215 AA.

XX AC AAR91044;

DT 23-MAY-1996 (first entry)

DE V8 mature protease (aal-215).

XX Linker peptide: V8 protease; Staphylococcus aureus;  
 KW recombinant protein; fusion protein; beta-galactosidase;  
 KW Escherichia coli; transposon Tn903;  
 KW aminoglycoside 3'-phosphotransferase.

XX Staphylococcus aureus strain V8 (ATCC 27733).

XX EP700995-A2.

XX 13-MAR-1996.

XX 06-SEP-1995; 95EP-0306235.

XX 07-NOV-1994; 94JP-0296028.

XX 07-SEP-1994; 94JP-0238595.

XX (SUNR ) SUNTORY LTD.

XX Ohsuye K, Yabuta M;

XX WPI; 1996-141021/15.

PT Prodn. of recombinant polypeptide(s) - using host cells transformed  
 PT with a gene coding for the desired polypeptide fused to a protective  
 PT polypeptide

PS Example 8; Fig 14c; 44pp; English.

XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A  
 CC and C are protective polypeptides, B is Staphylococcus aureus mature  
 CC V8 protease lacking the C-terminal repeat region, and L is a linker  
 CC (AAR91032). It was produced as an inclusion body in Escherichia coli  
 CC host cells, and was cleaved using the E. coli OmpT protease to  
 CC yield active V8 protease. Extension of the C-terminal end of the  
 CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)  
 CC i.e. up to residue Phe-215, also resulted in the formation of  
 CC inclusion bodies in E. coli. Any further extension gave a soluble  
 CC product which exhibited protease activity that repressed growth of  
 CC the host cells.

SQ Sequence 215 AA;

Alignment Scores:  
 Pred. No.: 0.541    Length: 215  
 Score: 10.00    Matches: 10  
 Percent Similarity: 100.00%    Conservative: 0  
 Best Local Similarity: 100.00%    Mismatches: 0  
 Query Match: 1.40%    Indels: 0  
 DB: 17    Gaps: 0

US-10-008-355-1 (1-2139) x AAR91044 (1-215)

QY 1930 ACGGCGGTAACTCCGGTAGCCCGGTATTC 1959  
 |||||  
 Db 165 ThrGlyGlyAsnSerGlySerProValphe 174

## RESULT 9

AAR29644  
 ID AAR29644 standard; Protein: 336 AA.

XX AC AAR29644;

DT 11-FEB-1993 (first entry)

DE Protease from S. Aureus.

DE Protease; PCR; amplify; Staphylococcus.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT Peptide 1..68

FT Protein 69..336

FT /label= Signal\_peptide

FT /label= Protease

XX JP04211370-A.

XX 03-AUG-1992.

XX 19-FEB-1991; 91JP-0024633.

XX 20-FEB-1990; 90JP-0040398.

XX (SHIO ) SHIONOGI & CO LTD.

XX WPI; 1992-304938/37.

XX N-PSDB; AAQ27988.

XX Novel protease prep'd. using Bacillus or Saccharomyces host -  
 PT capable of cleaving peptide bond at carboxyl terminus of glutamic  
 PT acid residues in polypeptide(s).

PS Disclosure; Page 15-16; 25pp; Japanese.

XX The sequences given in AAR26842 and AAR29644 are proteases which were  
 CC isolated from Staphylococcus aureus strains. The DNA sequences  
 CC encoding these proteins were isolated by PCR using the primer  
 CC sequences given in AAQ27960-86. The protease specifically cleaves  
 CC the peptide bond at the C-terminus of the glutamic acid residue in  
 CC polypeptide.

SQ Sequence 336 AA;

Alignment Scores:  
 Pred. No.: 0.519    Length: 336  
 Score: 10.00    Matches: 10  
 Percent Similarity: 100.00%    Conservative: 0  
 Best Local Similarity: 100.00%    Mismatches: 0  
 Query Match: 1.40%    Indels: 0  
 DB: 13    Gaps: 0

US-10-008-355-1 (1-2139) x AAR29644 (1-336)

QY 1930 ACGGCGGTAACTCCGGTAGCCCGGTATTC 1959  
 |||||  
 Db 233 ThrGlyGlyAsnSerGlySerProValphe 242

## RESULT 10

AAR91033  
 ID AAR91033 standard; Protein: 344 AA.

XX AC AAR91033;

XX 23-MAY-1996 (first entry)

XX Beta-galactosidase-V8 protease fusion protein.  
 DE V8 protease; Staphylococcus aureus; recombinant protein;  
 KW fusion protein; beta-galactosidase; Escherichia coli.  
 KW Chimeric Escherichia coli;  
 XX Chimeric Staphylococcus aureus strain V8 (ATCC 27733).  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..124  
 FT /note= "beta-galactosidase region"  
 FT Region 125..344  
 FT /note= "mature V8 protease without the repeat  
 FT region"  
 XX  
 PN EP700995-A2.  
 XX  
 PD 13-MAR-1996.  
 XX  
 XX 06-SEP-1995; 95EP-0306235.  
 PF  
 XX 07-NOV-1994; 94JP-0296028.  
 PR  
 XX 07-SEP-1994; 94JP-0238595.  
 XX  
 PA (SUNR ) SUNTORY LTD.  
 XX  
 XX Ohsuye K, Yabuta M;  
 PI WPI; 1996-141021/15.  
 DR

PT Prodn. of recombinant polypeptide(s) - using host cells transformed  
 PT with a gene coding for the desired polypeptide fused to a protective  
 PT polypeptide

PS Example 2; Page 12-13; 44pp; English.

XX Fusion proteins (AAR91033 and AAR91034) were constructed comprising  
 CC an Escherichia coli beta-galactosidase derivative (protective  
 CC polypeptide) fused at its C-terminal end to the S. aureus mature  
 CC V8 protease without or with the repeat region. The constructs  
 CC were inserted into vector pG7SADHCT(G)R6, yielding pV8RPT(-) and  
 CC pV8RPT(+), respectively. Both constructs yielded active protease  
 CC when expressed in E. coli JM101 transformants.  
 XX

SQ Sequence 344 AA;

Alignment Scores:  
 Pred. No.: 0.518 Length: 344  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 17 Gaps: 0

US-10-008-355-1 (1-2139) x AAR91033 (1-344)

QY 1930 ACGGGCGGTAACTCCGGTAGCCCGGTATTC 1959

Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 11

AAW22218

ID AAW22218 standard; Protein; 344 AA.

AC AAW22218;

XX 11-SEP-1997 (first entry)

XX Protein encoded by pV8RPT(-) construct.

DE Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;  
 KW truncation; wild type; PCR; polymerase chain reaction; amplification;  
 KW

XX proteolytic activity; fusion protein; beta-galactosidase; urea.  
 XX Chimeric - Escherichia coli.  
 OS Chimeric - Staphylococcus aureus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..100  
 FT /note= "E. coli beta-galactosidase portion"  
 FT Region 101..120  
 FT /note= "synthetic R6 linker"  
 FT Region 125..344  
 FT /note= "truncated S. aureus V8 protease portion"  
 XX  
 PN EP745669-A2.  
 XX  
 PD 04-DEC-1996.  
 XX  
 PF 31-MAY-1996; 96EP-0303939.  
 XX  
 PR 02-JUN-1995; 95JP-0170086.  
 XX  
 PA (SUNR ) SUNTORY LTD.  
 XX  
 XX Ohsuye K, Yabuta M;  
 PI WPI; 1997-013693/02.  
 DR  
 XX Staphylococcus aureus V8 protease mutants - with increased  
 PT resistance to denaturation  
 PT  
 XX Claim 2; Page 13-14; 42pp; English.

XX The invention relates to new mutant Staphylococcus aureus V8 proteases  
 CC which have enzyme activity even under environmental conditions which  
 CC promote protein denaturation. The mutants are based on 3 truncated V8  
 CC proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino  
 CC acids from the C-terminal of the wild type protease. The mutants also  
 CC contain amino acid substitutions, especially D44E, N71S and/or R147K.  
 CC The protein sequence shown here represents a chimaeric protein  
 CC comprising a truncated Staphylococcus aureus V8 protease lacking the  
 CC prepro and C-terminal 48 amino acids linked, via a synthetic linker,  
 CC downstream of the E. coli beta-galactosidase. The S. aureus portion of  
 CC the chimaera was amplified by the primers AAT73254-5 from wild type  
 CC sequence. The coding sequence was then used to generate mutants of the  
 CC V8 protease which retain their levels of activity in the presence of a  
 CC higher concentration of protein denaturant e.g. 5 M urea.  
 XX

SQ Sequence 344 AA;

Alignment Scores:  
 Pred. No.: 0.518 Length: 344  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 18 Gaps: 0

US-10-008-355-1 (1-2139) x AAW22218 (1-344)

QY 1930 ACGGGCGGTAACTCCGGTAGCCCGGTATTC 1959

Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298

RESULT 12

AAR26842

ID AAR26842 standard; Protein; 357 AA.

XX AAR26842;

XX 11-FEB-1993 (first entry)

XX Protease from S. Aureus ATCC12600.  
 DE  
 XX

03  
29

```

KW Protease; PCR; amplify; Staphylococcus.
XX Staphylococcus aureus.
OS
FH Key Location/Qualifiers
FT Peptide 1..68
FT Protein /label= Signal_peptide
FT 69..358
FT /label= Protease
XX
PN JP04211370-A.
XX
XX 03-AUG-1992.
XX
XX 19-FEB-1991; 91JP-0024633.
XX
XX 20-FEB-1990; 90JP-0040398.
XX
XX (SHIO ) SHIONOGI & CO LTD.
XX
XX WPI; 1992-304938/37.
XX
XX N-PSDB; AAQ27987.
XX
XX Novel protease prep'd. using Bacillus or Saccharomyces host -
XX capable of cleaving peptide bond at carboxyl terminus of glutamic
XX acid residues in polypeptide(s)
XX
XX Disclosure; Page 13-15; 25pp; Japanese.
XX
XX The sequences given in AAR26842 and AAR29644 are proteases which were
XX isolated from Staphylococcus aureus strains. The DNA sequences
XX encoding these proteins were isolated by PCR using the primer
XX sequences given in AAQ27960-86. The protease specifically cleaves
XX the peptide bond at the C-terminus of the glutamic acid residue in
XX polypeptide.
XX
XX Sequence 357 AA;
SQ
Alignment Scores:
Pred. No.: 0.516 Length: 357
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 13 Gaps: 0
US-10-008-355-1 (1-2139) x AAR26842 (1-357)
Oy 1930 ACGGGCGGTAACTCCGGTAGCCCGGTATTC 1959
Db 233 ThrGlyGlyAsnSerGlySerProValPhe 242
RESULT 13
AAR91034
ID AAR91034 standard; Protein; 392 AA.
XX
XX AAR91034;
AC
XX
XX 23-MAY-1996 (first entry)
DT
XX
XX Beta-galactosidase-V8 protease fusion protein.
DE
XX
XX V8 protease; Staphylococcus aureus; recombinant protein;
XX fusion protein; beta-galactosidase; Escherichia coli.
XX
XX Chimeric Escherichia coli;.
OS
XX Chimeric Staphylococcus aureus strain V8 (ATCC 27733).
XX
XX Key Location/Qualifiers
FH Region 1..124
FT /note= "beta-galactosidase region"
FT 125..392
FT /note= "mature V8 protease including the repeat
FT

```

---

```

FT XX region"
XX EP700995-A2.
PN
XX
XX 13-MAR-1996.
PD
XX
XX 06-SEP-1995; 95EP-0306235.
XX
XX 07-NOV-1994; 94JP-0296028.
XX
XX 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR ) SUNTORY LTD.
PA
XX
XX Ohsuye K, Yabuta M;
XX
XX WPI; 1996-141021/15.
XX
XX Prodn. of recombinant polypeptide(s) - using host cells transformed
XX with a gene coding for the desired polypeptide fused to a protective
XX polypeptide
XX
XX Example 2; Page 14-15; 44pp; English.
XX
XX Fusion proteins (AAR91033 and AAR91034) were constructed comprising
XX an Escherichia coli beta-galactosidase derivative (protective
XX polypeptide) fused at its C-terminal end to the S. aureus mature
XX V8 protease without or with the repeat region. The constructs
XX were inserted into vector pG7S4Dhct(G)R6, yielding pV8RPT(-) and
XX pV8RPT(+), respectively. Both constructs yielded active protease
XX when expressed in E. coli JM101 transformants.
XX
XX Sequence 392 AA;
SQ
Alignment Scores:
Pred. No.: 0.512 Length: 392
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 17 Gaps: 0
US-10-008-355-1 (1-2139) x AAR91034 (1-392)
Oy 1930 ACGGGCGGTAACTCCGGTAGCCCGGTATTC 1959
Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298
RESULT 14
AAR91035
ID AAR91035 standard; Protein; 532 AA.
XX
XX AAR91035;
AC
XX
XX 23-MAY-1996 (first entry)
DT
XX
XX Recombinant V8 protease V8D fusion protein.
DE
XX
XX Linker peptide; V8 protease; Staphylococcus aureus;
XX recombinant protein; fusion protein; beta-galactosidase;
XX Escherichia coli; transposon Tn903;
XX aminoglycoside 3'-phosphotransferase.
XX
XX Chimeric Escherichia coli;
OS
XX Chimeric synthetic;
OS
XX Chimeric Staphylococcus aureus strain V8 (ATCC 27733);
OS
XX Chimeric transposon Tn903.
XX
XX Key Location/Qualifiers
FH Region 1..100
FT /note= "beta-galactosidase region"
FT 101..120
FT /note= "R6 linker"
FT 104..105
FT Cleavage-site
FT

```

FT Region /note= "cleavage site for OmpT protease"  
 FT 125..335  
 FT /note= "V8 protease region"  
 FT 336..356  
 FT /note= "R6 linker"  
 FT 339..340  
 FT Cleavage-site  
 FT /note= "cleavage site for OmpT protease"  
 FT 307..532  
 FT /note= "aminoglycoside 3'-phosphotransferase region"

FT EP700995-A2.  
 FT 13-MAR-1996.  
 XX 06-SEP-1995; 95EP-0306235.  
 XX 07-NOV-1994; 94JP-0296028.  
 PR 07-SEP-1994; 94JP-0238595.  
 XX (SUNR ) SUNTORY LTD.  
 PA Ohsuye K, Yabuta M;  
 PI WPI; 1996-141021/15.  
 DR

XX Prodn. of recombinant polypeptide(s) - using host cells transformed with a gene coding for the desired polypeptide fused to a protective polypeptide  
 XX Example 3; Page 16-18; 44pp; English.

XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A and C are protective polypeptides (derived from Escherichia coli beta-galactosidase and and Tn903 aminoglycoside 3'-phosphotransferase, respectively), B is Staphylococcus aureus mature V8 protease lacking the C-terminal repeat region, and L is a linker peptide (AAR91032). The fusion protein is expressed in inactive form in E. coli. It is then recovered, solubilised and cleaved at the linker peptide regions with a protease intrinsic to the host cells, i.e. OmpT protease, to allow recovery of V8 protease.

XX Sequence 532 AA;

Alignment Scores:  
 Pred. No.: 0.498 Length: 532  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 17 Gaps: 0

US-10-008-355-1 (1-2139) x AAR91035 (1-532)

QY 1930 ACGGCGGTAACTCCGTTAGCCCGTATTC 1959  
 Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298  
 ||||||||||||||||||||||||||||||||||

RESULT 15  
 AAW22219  
 ID AAW22219 standard; Protein: 532 AA.  
 AC AAW22219;

XX 11-SEP-1997 (first entry)  
 AC Protein encoded by pV8D construct.

XX Mutant: Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;  
 KW truncation; wild type; PCR; polymerase chain reaction; amplification;  
 KW proteolytic activity; fusion protein; beta-galactosidase; urea.  
 XX Chimeric - Escherichia coli.

OS Chimeric - Staphylococcus aureus.  
 XX Key Location/Qualifiers  
 FH Region 1..100  
 FT /note= "E. coli beta-galactosidase portion"  
 FT 101..124  
 FT /note= "R6 linker sequence"  
 FT 125..336  
 FT /note= "truncated S. aureus V8 protease portion"  
 FT 337..360  
 FT /note= "R6 linker sequence"  
 FT 361..532  
 FT /note= "aminoglycoside 3'-phosphotransferase portion"

XX EP745669-A2.  
 PN 04-DEC-1996.  
 PD 31-MAY-1996; 96EP-0303939.  
 XX 02-JUN-1995; 95JP-0170086.  
 PR (SUNR ) SUNTORY LTD.  
 PA Ohsuye K, Yabuta M;  
 PI WPI; 1997-013693/02.  
 DR Staphylococcus aureus V8 protease mutants - with increased resistance to denaturation  
 XX Claim 7; Page 16-17; 42pp; English.

XX The invention relates to new mutant Staphylococcus aureus V8 proteases which have enzyme activity even under environmental conditions which promote protein denaturation. The mutants are based on 3 truncated V8 proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino acids from the C-terminal of the wild type protease. The mutants also contain amino acid substitutions, especially D44E, N71S and/or R147K. The protein sequence shown here represents a chimaeric protein comprising a truncated Staphylococcus aureus V8 protease lacking the prepro and C-terminal 56 amino acids linked, via a synthetic R6 linker, downstream of the E. coli beta-galactosidase. Also included downstream of the V8 protease fragment is a second R6 linker and a fragment of the aminoglycoside 3'-phosphotransferase protein. The chimaeric sequence was generated by restriction digestion and ligation from the V8RPT(-) sequence (see AAW22218) by using a natural EcoRV site which removed a further 8 amino acids from the C-terminus. This truncated V8 protease, designated V8D, retains its level of activity in the presence of a higher concentration of protein denaturant e.g. 5 M urea.

XX Sequence 532 AA;

Alignment Scores:  
 Pred. No.: 0.498 Length: 532  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 18 Gaps: 0

US-10-008-355-1 (1-2139) x AAW22219 (1-532)

QY 1930 ACGGCGGTAACTCCGTTAGCCCGTATTC 1959  
 Db 289 ThrGlyGlyAsnSerGlySerProValPhe 298  
 ||||||||||||||||||||||||||||||||||

Search completed: May 16, 2003, 13:06:34  
 Job time : 123.5 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 13:13:37 ; Search time 83 Seconds  
(without alignments)  
4971.283 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 712

Sequence: 1 atgcaaatgaaatataaaag.....aagagctgaattgatctaa 2139

Scoring table:

OLIGO Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 362588 seqs, 96450795 residues

Word size: 1

Total number of hits satisfying chosen parameters: 683032

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US10008355/runat\_16052003\_110429\_9845/app\_query.fasta\_1.2311  
-DB=PublishedApplications\_AA -QFMT=fastan -SUFFIX=olin2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1  
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10008355 -CGN\_1\_1\_25 @runat\_16052003\_110429\_9845  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELEX=7

Database :

PublishedApplications\_AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	712	100.0	712	9	US-10-008-355-2
2	699	98.2	699	9	US-10-008-355-8
3	52	7.3	52	9	US-10-008-355-3
4	21	2.9	26	9	US-10-008-355-18

Alignment Scores:

Pred. No.: 0  
Score: 712.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
Length: 712  
Matches: 712  
Conservative: 0  
Mismatches: 0  
Indels: 0

ALIGNMENTS

RESULT 1

US-10-008-355-2

; Sequence 2, Application us/10008355

; Patent No. US20020164759A1

; GENERAL INFORMATION:

; APPLICANT: Travis, James

; APPLICANT: Potempa, Jan S

; APPLICANT: Barbula, Agnieszka

; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101

; CURRENT APPLICATION NUMBER: US/10/008.355

; CURRENT FILING DATE: 2001-11-08

; PRIOR APPLICATION NUMBER: US 60/246,827

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 712

; TYPE: PRT

; ORGANISM: Porphyromonas gingivalis

US-10-008-355-2

Sequence 7, Appli  
Sequence 5, Appli  
Sequence 25, Appl  
Sequence 4, Appli  
Sequence 6, Appli  
Sequence 26, Appl  
Sequence 9, Appli  
Sequence 30, Appl  
Sequence 17, Appl  
Sequence 51, Appl  
Sequence 10510, A  
Sequence 44, Appl  
Sequence 18, Appl  
Sequence 3, Appli  
Sequence 2, Appli  
Sequence 4, Appli  
Sequence 89, Appl  
Sequence 36591, A  
Sequence 38461, A  
Sequence 57, Appl  
Sequence 39870, A  
Sequence 39443, A  
Sequence 266, App  
Sequence 368, App  
Sequence 16, Appl  
Sequence 1089, Ap  
Sequence 324, App  
Sequence 324, App  
Sequence 5217, Ap  
Sequence 1813, Ap  
Sequence 1970, Ap  
Sequence 1441, Ap  
Sequence 5670, Ap  
Sequence 1694, Ap  
Sequence 961, App  
Sequence 29, Appl  
Sequence 29, Appl  
Sequence 154, App  
Sequence 44523, A  
Sequence 130, App

DB: 9 Gaps: 0  
US-10-008-355-1 (1-2139) x US-10-008-355-2 (1-712)  
QY 1 ATGCAGATGAATTAAGATGTTCTTCGAGCAGCCCTGCTGGGTGCTTCAGGG 60  
DB 1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuGlyAlaSerGly 20  
QY 61 GTAGCCAAAGCCGCAAGGCGATGGCTCTCAACGAACTCAATCAGGAGAATCTGGAT 120  
DB 21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlnLeuAsnLeuAsp 40  
QY 121 CGAATGCGTGAGCTGGCTTTACGCTCCGCTGGATTGCTCTACAGTTTCACAAAGCCG 180  
DB 41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60  
QY 181 TCCATTGCCAATGCCGTGGTTATCTTCGGTGGCGGATGATCCGGTATCACAGTTCGGAT 240  
DB 61 SerIleAlaAsnAlaValAlaPheGlyGlyCysThrGlyIleThrValSerAsp 80  
QY 241 CAGGCGCTGATCTTACCACACACCTGCGGATACGGTGCTATCCAGAGCCAAAGCACG 300  
DB 81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100  
QY 301 GTGGATCACGACTATCTCCGCGATGGTTTCGTTCTCCGACGATGGGTGAGAGCTTCG 360  
DB 101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluLeuPro 120  
QY 361 ATTCGGGTCTTTCCGTCGAAGTATCTGCGCAAGATCTGGAAGTTAACGGCAAGGTAGAA 420  
DB 121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140  
QY 421 GGACAGCTCAAGGGTATCACGTGACGAGATGGAGCTCTCGCCAAAGCTCAGGAGTATGC 480  
DB 141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnLysValCys 160  
QY 481 CAAGAACTGGCCAAAGAAATGACAGACGAGAACCACTCTGCGATCGTAGAGCTTTC 540  
DB 161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180  
QY 541 TATTCACAACGAACTCTCTCATCTCTACGATGATTCACAGGAGCTTCGTATGGTA 600  
DB 181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200  
QY 601 TTTGCTCTCCAGCTCTGTAGTAAAGTTTCGAGGCGATACGCAACTGGATGTGGCCG 660  
DB 201 PheAlaProSerSerValGlyLysPheGlyAspThrAspAsnTrpMetTrpPro 220  
QY 661 CGTCACGCGGCTTACGCGTATTCGCGGTATGCGGTGCGGACACCGCGCGCC 720  
DB 221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240  
QY 721 GAATACACAAGGCAATTAACCCCTATAAGCCGTTTACTTCGCTGCGGTATCCATGCAA 780  
DB 241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260  
QY 781 GGCTACAAAGGTGAGCACTATGCCATGACCTCGGTTTCCGCGGCGAGTACGATCGCTAC 840  
DB 261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280  
QY 841 CTCACCTTCTGGGTGCGAGATCGTATCGAAGAACGAGACAACTCTCGTATCGAAGTT 900  
DB 281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGlnAsnAsnProArgIleGluVal 300  
QY 901 CGCGTATCAAGAGGATCTGGAAGAACCATGAGCGAGATCAGGCTACCGCTATCCGATC 960  
DB 301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320  
QY 961 AAATATGCCAGCAAGTATGCTCAGAGTCTAACTATTCGGAAGAATTCGATCGGTATGAC 1020  
DB 321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340  
QY 1021 CGCGTCTCGCTCTGAGCTGATAGCTGTAAGCGTGGCGGAGGAAGACCATTCGCA 1080

DB 341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360  
QY 1081 GACTCGGATCCGTAAGAACGCGCAAGAGTCTCTATGCGGATGATTGCTCTCTCTCGAA 1140  
DB 361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380  
QY 1141 AAGCTTTATAAGGAGGAGGCCAAGCCGTCAGACTGACTTATTTGAGGAGACGCTC 1200  
DB 381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400  
QY 1201 TTCGGTGGTACCGAGGTGGTTCTGTTTGCACAGTTTGCACACGCTATCGCTACAAATCCT 1260  
DB 401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420  
QY 1261 GATCTCATGCTGCGGTATCTCAATCGCTTACGACAGTACAAAGACTACTCCCTCG 1320  
DB 421 AspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSer 440  
QY 1321 CTCGACCGTAAGGTGCTGCCCGCATGCTCGATATGCTACGCCGCGGTATCCCTCGCGAC 1380  
DB 441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp 460  
QY 1381 AAGCTCCCGATATATTCAAGAATGTAATCGACAGAATAATCAAGGCGGACACCAAGAG 1440  
DB 461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480  
QY 1441 TATCAGACTTCGTATTCGACAAGAGTGTGCTTCTATAGCCACAGTTCATCCGCTCCATG 1500  
DB 481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500  
QY 1501 CTCAGTTCATGGACAAGGAAAGCTTTGCCAAGCTATTCGAGAAGATCCGCGCATAGAG 1560  
DB 501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520  
QY 1561 CTTTCCAGACCGTAATAGTCTGCTCGCGCTATTACGCGCGGTATCGCATGCCAAATGCC 1620  
DB 521 LeuSerLysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540  
QY 1621 TATGCCATTGAGAAGGCAAGCGCTTTCTTCTGCGGTTTGGCTGAGATGATACCCCGGA 1680  
DB 541 TyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGluMetTyrProGly 560  
QY 1681 CGTCTCTCGCGAGCGATGCCAATCTCACATGCGTATGAGCTACGCGTCCATCAAGGGA 1740  
DB 561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580  
QY 1741 TATGACCGCAGGAGCGTCTGCTGCTACAACTATCATACGACAGCAAGGCGGTATGGAG 1800  
DB 581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600  
QY 1801 AAGCAGGATCCTTAAGAGCGATGAGTTTCCGCTACAGGAGAATATCTCGACTCTTCGCG 1860  
DB 601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620  
QY 1861 ACCAAAACTATGTCGCTATGCGGAGAACCGTCAGCTCCATATCGCTTCTCTTCATCGAAC 1920  
DB 621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640  
QY 1921 AACGACATCACGGCGGTAACTCCGGTAGCCCGCTATTTCGATAAGACGCGCGTCTGATC 1980  
DB 641 AsnAspIleThrGlyLysSerGlySerProValPheAspLysAsnGlyArgLeuIle 660  
QY 1981 GGTCTTGTCTTCGATGGCAACTGGGAAGCTATGATGTGTGATCGAGTTCGAGTTCGAACCCGAT 2040  
DB 661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680  
QY 2041 CTGAGCGCCACAATCAGCGTGGACATCCGCTACGTTCTCTTCATGATGATGACAAATGGGT 2100  
DB 681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700  
QY 2101 CAGTCCCGCTCTCATCCAAAGACTGAAGTTGATC 2136  
|||||

Db 701.GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712

## RESULT 2

US-10-008-355-8  
; Sequence 8, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 699  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-8

## Alignment Scores:

Pred. No.: 0 Length: 699  
Score: 699.00 Matches: 699  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.17% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-10-008-355-8 (1-699)

QY	1	ATCAAAATGAAATTAATAAGATATCTTCGAGCAGCCCTGCTGTTGGGTTCAGGG	60
Db	1	MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuGlyAlaSerGly	20
QY	61	GTAGCCAAAGCCGACAAAGCATGTGCTCTCAACGAACCTCAATCAGGAGAATCTGGAT	120
Db	21	ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlnLeuAsnLeuAsp	40
QY	121	CGAATCGTGAGCTCGGCTTTACGCTCCGTTGGATTGCTCTACATGTTTCGACAAAGCCG	180
Db	41	ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro	60
QY	181	TCCATTGCCAATGCCGNGGTATCTTCGTTGGCGGATGATCCGCTATCAGAGTCTCCGAT	240
Db	61	SerIleAlaAsnAlaValAlaValIlePheGlyGlyCysThrGlyIleThrValSerAsp	80
QY	241	CAGGCGCTGATCTTTAGCAACCAACCACTGCGGATACGGTACGTATCCAGAGCCAAAGCAGC	300
Db	81	GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr	100
QY	301	GTGATCAGCAGTATCTGCGGATGGTTTCGTTCTCGCAGATGGGTGAGGAGCTTCGG	360
Db	101	ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluLeuPro	120
QY	361	ATTCGCGGTCTTTCCGCGAAGTATCTCGCAAGATCTGGAAGTTAAACGACCAAGGTAGAA	420
Db	121	IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu	140
QY	421	GGACAGCTCAAGGTATCACTGACGAGATGGAGGTCTGCGCAAGCTCAGGAGGTATGC	480
Db	141	GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys	160
QY	481	CAAGAATCGCCAAAAGAAATGACAGACGAGAACCAACTCTGCACTCGTAGAGCCTTTC	540
Db	161	GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe	180
QY	541	TATTCACACAGCAATACTCTCCCTCATCGTCTACCATGATTTCAGGAGCTTCGTATGGTA	600
Db	181	TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal	200

QY	601	TTTGTCTCTCCAGCTCTGTAGGTAAGTTCCGAGCCGATACGGACAACACTGGATCGCCG	660
Db	201	PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro	220
QY	661	CGTCACACGGCGACTTCACGGTATCCGGGTGTATCCGGTCCGACACACCGCCGCGCC	720
Db	221	ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla	240
QY	721	GAATACACCAAGGACAAATAAACCCCTATAAGCCCTTTACTTCGCTGCGGTATCCATCAA	780
Db	241	GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln	260
QY	781	GGCTACAAAGCTGACGACTATGCCATGACCATCGGTTTCCCGGCAGTACGATCGGTAC	840
Db	261	GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr	280
QY	841	CTCACTTCTTGGGCTGTGGAAGATCGTATCGAAACGAGAACAAATCTCGTATCGAAGTT	900
Db	281	LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal	300
QY	901	CGCGGTATCAAGCAAGGCATCTGGAAGAACCCATGAGCCGACATCAGGCTACCGGTATC	960
Db	301	ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle	320
QY	961	AAATATGCCAGCAAGTATGCTCAGAGTGCCTAACTATTGGAAGAATTTCGATCGGTATGAA	1020
Db	321	LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn	340
QY	1021	CGCGGTCTCGCTCGTTCGACGTATAGTGTGTAAGCGTCCGAGGAAAGAGCATTCGCA	1080
Db	341	ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluAlaArgAlaPheAla	360
QY	1081	GACTGGATCCGTAGAACGCGCAAGTGCCTCTATGCGGATGTATTGCTCTCTCTCGAA	1140
Db	361	AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu	380
QY	1141	AAGCTTTATAGGAAGGAGCCAAAGCCGTCAGATGACTTATTTGAGCGAGACGCTC	1200
Db	381	LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu	400
QY	1201	TTCGGTGGTACCGAGGTGGTTCGTTTTCGACAGTTCGCCAACGATGGCTACAAATCCT	1260
Db	401	PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro	420
QY	1261	GATCCTCATGCCGTATCCTCAAAATCGCTGACGACAGTACAAAGACTACCTCCCTCG	1320
Db	421	AspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSer	440
QY	1321	CTCGACCGTAAAGTGCCTCCCGCCATGCTCGATATTGTCGCGCGCTATCCCTGCGGAC	1380
Db	441	LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp	460
QY	1381	AAGTCCCGCATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAG	1440
Db	461	LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys	480
QY	1441	TATCAGACTTCGTATTCGACAAGTGTGTTCTTATACGACAGTTCATCCATCCCATG	1500
Db	481	TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet	500
QY	1501	CTCAAGTCCATGGCAAGGAAAGTTTCCCAAGGCTATCGAGAAGATCCCGCAGTAGAC	1560
Db	501	LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu	520
QY	1561	CTTTCCAAAGCGTAATAGCTGCTGCGCTATTCAGGCGCGATCGGATGCCAATGCC	1620
Db	521	LeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla	540
QY	1621	TATCCCATGGAAGGCGCAAGCGCTCTTTCTTTCGCGGTTCGCTGAGATGTACCCCGGA	1680
Db	541	TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly	560

```
Qy 1681 CGTCTCTGCCAGCGATGCGCAACTTACCATGCGTATGAGTACGGCTCCATCAAGGA 1740
Db 561 ArgAlaLeuProSerAspAlaAsnPhetThrMetArgMetSerTyrGlySerIleLysGly 580
Qy 1741 TATGAACCGCAGGAGCGTGCCTGGTACAACTATCATAGCAGCAAGGGCGTATTGGAG 1800
Db 581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600
Qy 1801 AAGCAGGATCTTAAGAGCGATGAGTTGCTCCGTACAGGAGATATCTCGACCTCTCCGC 1860
Db 601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620
Qy 1861 ACCAAACTATGTTCGCTATGCCAGAACGGTCAGCTCCATATCGCTTCTTATCGAAC 1920
Db 621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640
Qy 1921 AACGACATCAGCGCGGTAACTCCGGTAGCCCGGTATTCGATAAGACGGCGCTCTGATC 1980
Db 641 AsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660
Qy 1981 GGTCTTGTCTTGGATGGCAACTGGGAAGCTATGAGTGTGATCGAGTTCGAACCCGAT 2040
Db 661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680
Qy 2041 CTGAGCGCACAAATCAGCGTGGACATCGGTCAGCTTCTTTCATGATTGACAAATGG 2097
Db 681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrp 699

RESULT 3
US-10-008-355-3
; Sequence 3, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-3
Alignment Scores:
Pred. No.: 5,6e-42 Length: 52
Score: 52,00 Matches: 52
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 7,30% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-10-008-355-3 (1-52)
Qy 1930 ACGGGCGGTAACTCCGGTAGCCCGTATTCGATAAGAACGCCGCTGTGATCGGCTTGCT 1989
Db 1 ThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAla 20
Qy 1990 TTCATGCGCAACTGGGAGCTATGAGTGTGATCGAGTTCGACATCGAGTTCGAACCCGATCTCAGCGC 2049
Db 21 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArg 40
Qy 2050 ACAATCAGCGTGGACATCCGCTAGCTTCTCTTCATG 2085
Db 41 ThrIleSerValAspIleArgTyrValLeuPheMet 52

RESULT 4
```

```
US-10-008-355-18
; Sequence 18, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-18
Alignment Scores:
Pred. No.: 1,86e-11 Length: 26
Score: 21,00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 2,95% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-10-008-355-18 (1-26)
Qy 82 ATGTGGTCTCTCAACGAACCTCAATCAGAGAACTGTGGATCGAATCGGTAGCTCGGCTTT 141
Db 6 MetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPhe 25
Qy 142 ACG 144
Db 26 Thr 26

RESULT 5
US-10-008-355-7
; Sequence 7, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-008-355-7
Alignment Scores:
Pred. No.: 0,000938 Length: 716
Score: 13,00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1,83% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-10-008-355-7 (1-716)
Qy 643 GACAACCTGGATGTGGCCGCGTCACACGGCGCAGCTTCAGC 681
Db 1
```

Db 212 AspAsnTrpMetTrpProArgHisThrGlyAspPheSer 224  
RESULT 6  
US-10-008-355-5  
; Sequence 5, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 734  
; TYPE: PRT  
; ORGANISM: Shewanella putrefaciens  
US-10-008-355-5  
Alignment Scores:  
Pred. No.: 0.0867 Length: 734  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.54% Indels: 0  
DB: 9 Gaps: 0  
US-10-008-355-1 (1-2139) x US-10-008-355-5 (1-734)  
QY 643 GACACGTGGATGGCGCGCTCACACGGCGCAC 675  
Db 222 AspAsnTrpMetTrpProArgHisThrGlyAsp 232  
RESULT 7  
US-10-008-355-25  
; Sequence 25, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 25  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus sequence  
US-10-008-355-25  
Alignment Scores:  
Pred. No.: 1.37 Length: 10  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 9 Gaps: 0  
US-10-008-355-1 (1-2139) x US-10-008-355-25 (1-10)  
QY 1930 ACGGGCGGTAACTCCGGTAGCCCGGTATTC 1959

Db 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
RESULT 8  
US-10-008-355-4  
; Sequence 4, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-008-355-4  
Alignment Scores:  
Pred. No.: 1.13 Length: 52  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 9 Gaps: 0  
US-10-008-355-1 (1-2139) x US-10-008-355-4 (1-52)  
QY 1930 ACGGGCGGTAACTCCGGTAGCCCGGTATTC 1959  
Db 1 ThrGlyGlyAsnSerGlySerProValPhe 10  
RESULT 9  
US-10-008-355-6  
; Sequence 6, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 732  
; TYPE: PRT  
; ORGANISM: Shewanella putrefaciens  
US-10-008-355-6  
Alignment Scores:  
Pred. No.: 0.835 Length: 732  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 9 Gaps: 0  
US-10-008-355-1 (1-2139) x US-10-008-355-6 (1-732)  
QY 1930 ACGGGCGGTAACTCCGGTAGCCCGGTATTC 1959

Db 662 ThrGlyGlyAsnSerGlySerProValPhe 671

## RESULT 10

US-10-008-355-26  
; Sequence 26, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 26  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus sequence  
US-10-008-355-26

Alignment Scores:  
Pred. No.: 2.14e+07 Length: 9  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.26% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-10-008-355-26 (1-9)

Qy 1930 ACGGCGGTAACTCCGGTAGCCCGTA 1956

Db 1 ThrGlyGlyAsnSerGlySerProVal 9

## RESULT 11

US-10-008-355-9  
; Sequence 9, Application US/10008355  
; Patent No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235.00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-10-008-355-9

Alignment Scores:  
Pred. No.: 8.05 Length: 720  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.26% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-10-008-355-9 (1-720)

Qy 655 TGGCCGGTACACGGCGGACTTCAGC 681

Db 221 TrpProArgHisThrGlyAspPheSer 229

## RESULT 12

US-10-166-087-30  
; Sequence 30, Application US/10166087  
; Publication No. US2003007767A1  
; GENERAL INFORMATION:  
; APPLICANT: Ecopia Biosciences Inc.  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: Genes and proteins for the biosynthesis of anthramycin  
; FILE REFERENCE: 3014-20S  
; CURRENT APPLICATION NUMBER: US/10/166,087  
; CURRENT FILING DATE: 2002-06-11  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 30  
; LENGTH: 274  
; TYPE: PRT  
; ORGANISM: Streptomyces refuineus subspecies thermotolerans  
US-10-166-087-30

Alignment Scores:  
Pred. No.: 86.6 Length: 274  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-10-166-087-30 (1-274)

Qy 121 CGAATGCGTGAGCTCGGCTTTACG 144

Db 168 ArgMetArgGluLeuGlyPheThr 175

## RESULT 13

US-09-510-332-17  
; Sequence 17, Application US/09510332  
; Publication No. US20030022278A1  
; GENERAL INFORMATION:  
; APPLICANT: Zuker, Charles S.  
; APPLICANT: Adler, Jon Elliot  
; APPLICANT: Ryba, Nick  
; APPLICANT: Mueller, Ken  
; APPLICANT: Hoon, Mark  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: T2R, a No. US20030022278A1 Family of Taste Receptors  
; FILE REFERENCE: 02307E-098010US  
; CURRENT APPLICATION NUMBER: US/09/510,332  
; CURRENT FILING DATE: 2000-02-22  
; PRIOR APPLICATION NUMBER: US 09/393,634  
; PRIOR FILING DATE: 1999-09-10  
; NUMBER OF SEQ ID NOS: 172  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 312  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human T2R09 (hGR09)  
US-09-510-332-17

Alignment Scores:  
Pred. No.: 85.3 Length: 312  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.13% Indels: 0  
DB: 9 Gaps: 0

US-10-008-355-1 (1-2139) x US-09-510-332-17 (1-312)

Qy 1164 CTTGGCTCTTCTTATAAGCCTT 1141  
|||||  
Db 135 LeuGlySerPheLeuIleSerLeu 142

## RESULT 14

US-09-393-634-51  
; Sequence 51, Application US/09393634  
; Patent No. US20020051997A1  
; GENERAL INFORMATION:  
; APPLICANT: Zuker, Charles S.  
; APPLICANT: Adler, Jon Elliot  
; APPLICANT: Ryba, Nick  
; APPLICANT: Mueller, Ken  
; APPLICANT: Hoon, Mark  
; APPLICANT: The Regents of the University of California  
; APPLICANT: The Government of the United States of America  
; APPLICANT: as represented by the Secretary of the  
; APPLICANT: Department of Health and Human Services  
; TITLE OF INVENTION: SF, a No. US20020051997A1e1 Family of Taste Receptors  
; FILE REFERENCE: 02307E-098000US  
; CURRENT APPLICATION NUMBER: US/09/393,634  
; CURRENT FILING DATE: 1999-09-10  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 51  
; LENGTH: 312  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: human GR09  
US-09-393-634-51

Alignment Scores:  
Pred. No.: 85.3 Length: 312  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.13% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-1 (1-2139) x US-09-393-634-51 (1-312)

Qy 1164 CTTGGCTCTTCTTATAAGCCTT 1141  
|||||  
Db 135 LeuGlySerPheLeuIleSerLeu 142

## RESULT 15

US-09-815-242-10510  
; Sequence 10510, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10510  
; LENGTH: 659  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-815-242-10510

Alignment Scores:  
Pred. No.: 78.3 Length: 659  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: 10 Gaps: 0

US-10-008-355-1 (1-2139) x US-09-815-242-10510 (1-659)

Qy 490 GCCAAAAAGAAATGCAGACGAG 513  
|||||  
Db 206 AlaLysLysGluAsnAlaAspGlu 213

Search completed: May 16, 2003, 13:41:27  
Job time : 98 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 13:00:32 ; Search time 79 seconds  
(without alignments)  
5205.861 Million cell updates/sec

Title: us-10-008-355-1

Perfect score: 712

Sequence: 1 atgcaaatgaataaaag.....aagagctgaagtgtctaa 2139

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delopt 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565536

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlx  
-O=/cgn2.1/USPTO.spool/US10008355/runat\_16052003.110426.9746/app\_query.fasta\_1.2311  
-DB=PIR\_73 -OFMT=fastan -SUFFIX=olin2p.rpr -MINMATCH=0.1 -LOOPCL=0 -ICPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pt  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355.ecgn\_1.1.111.runat\_16052003.110426.9746 -ICPU=3  
-NO\_XLPHY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAEXT=60 -FGAPOP=6 -FGAEXT=7  
-YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database :  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	13	1.8	716	2 G82627	hypothetical prote
2	10	1.4	336	1 PRSASK	glutamyl endopepti
3	10	1.4	342	2 G89873	hypothetical prote
4	10	1.4	357	2 S21758	glutamic acid-spec
5	10	1.4	569	2 B75381	probable two-compo
6	9	1.3	266	2 D75483	hypothetical prote
7	9	1.3	452	2 AE1710	hypothetical prote
8	9	1.3	483	2 AF2363	hypothetical prote
9	9	1.3	684	2 T36771	probable integral
10	9	1.3	984	1 A34076	protein-tyrosine k
11	9	1.3	1116	2 B70476	hypothetical prote
12	8	1.1	102	2 T36095	hypothetical prote
13	8	1.1	106	2 T04804	hypothetical prote
14	8	1.1	116	2 JC4669	cocaine-and amphet

c	15	8	1.1	162	2 C69495	conserved hypothet
	16	8	1.1	189	2 A69054	hypothetical prote
	17	8	1.1	214	2 C84259	triosephosphate is
c	18	8	1.1	224	2 A81449	probable molybdenu
	19	8	1.1	226	2 S04752	H+-transporting tw
c	20	8	1.1	241	2 A69299	ABC transporter, A
	21	8	1.1	256	2 A34770	ORF1 protein - sal
	22	8	1.1	284	2 T07614	homeobox-leucine z
	23	8	1.1	288	2 F72219	conserved hypothet
c	24	8	1.1	332	2 S46947	ribosomal protein
	25	8	1.1	349	2 S40718	hypothetical prote
	26	8	1.1	382	2 T39451	methionine adenosy
	27	8	1.1	392	2 E83899	thiolase (acetyl-C
	28	8	1.1	400	2 E96029	probable beta-keto
	29	8	1.1	401	2 AI3115	beta-ketoadipyl Co
	30	8	1.1	419	2 D98171	beta-ketoadipyl-Co
c	31	8	1.1	421	2 C91109	integrase (importe
	32	8	1.1	421	2 F85954	probable pathogen
c	33	8	1.1	425	2 S54010	1-aminocyclopropan
	34	8	1.1	444	2 S54011	1-aminocyclopropan
	35	8	1.1	458	2 F75367	potassium uptake p
c	36	8	1.1	479	2 T44326	hypothetical prote
	37	8	1.1	513	2 T34689	probable iron-sulf
c	38	8	1.1	544	2 D64882	periplasmic oligop
	39	8	1.1	544	2 G90867	probable transport
	40	8	1.1	544	2 B85751	probable transport
	41	8	1.1	563	2 B97319	homolog of plant a
c	42	8	1.1	570	2 A48836	fibropellin C prec
	43	8	1.1	586	2 S71235	chaperonin 60 alph
c	44	8	1.1	633	2 G95385	Kup2 Potassium upt
	45	8	1.1	637	1 YCRP	acetolactate synth

ALIGNMENTS

RESULT 1

G82627  
hypothetical protein Xf1887 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:accession: G82627  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:accession: G82627  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-716 <SIM>  
A:Cross-references: GB:AE004008; GB:AE003849; NID:g9106961; PIDN:AAF84693.1; GSPDB:GN  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.P.; Marino, C.L.; Marques, M.V.; Martins  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Miyaki, C.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: Xf1887

Alignment Scores:  
Pred. No.: 0.000396  
Score: 13.00  
Percent Similarity: 100.00%  
Length: 716  
Matches: 13  
Conservative: 0

3A 102a  
100

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.83% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x G82627 (1-716)

Qy 643 GACAACTGGATGTCGCGGTACACGCGGCGACTTCAGC 681  
|||||  
Db 212 AspAsnTrpMetTrpProArgHisThrGlyAspPheSer 224

## RESULT 2

PRASASK  
glutamyl endopeptidase (EC 3.4.21.19) precursor - Staphylococcus aureus  
N:Alternate names: staphylococcal serine proteinase

C:Species: Staphylococcus aureus  
C:Date: 04-Dec-1986 #sequence\_revision 30-Jun-1991 #text\_change 18-Jun-1999

C:Accession: A26812; A00966

R:Carmona, C.; Gray, G.L.

Nucleic Acids Res. 15, 6757, 1987

A:Title: Nucleotide sequence of the serine protease gene of Staphylococcus aureus, strain

A:Reference number: A26812; MUID:87316953; PMID:3306605

A:Accession: A26812

A:Molecule type: DNA

A:Residues: 1-336 <CARG>

A:Cross-references: EMBL:Y00356; NID:946686; PIDN:CAA68434.1; PID:g46687

A:Experimental source: strain V8

R:Drapeau, G.R.

Can. J. Biochem. 56, 534-544, 1978

A:Title: The primary structure of staphylococcal protease.

A:Reference number: A23824; MUID:78212487; PMID:96922

A:Accession: A00966

A:Molecule type: protein

A:Residues: 69-108; 110-124; 126-144, 'D', 146-192, 'T', 194-228, 'N', 230-258, 'Q', 260, 'D', 262-2

A:Experimental source: strain V8

C:Comment: This extracellular proteolytic enzyme preferentially cleaves the peptide bond  
C:Comment: This enzyme may be distantly related to the trypsin-type serine proteinases,  
dues.

C:Superfamily: staphylococcal serine proteinase

C:Keywords: hydrolase; serine proteinase

F:69-336/Product: staphylococcal serine proteinase #status experimental <ACT>

F:119,161/Active site: His, Asp #status predicted

F:237/Active site: Ser #status experimental

## Alignment Scores:

Pred. No.: 0.513 Length: 336  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x PRASASK (1-336)

Qy 1930 ACGGCGGTAACTCGGTAGCCCGTATTC 1959  
|||||  
Db 233 ThrGlyGlyAsnSerGlySerProValPhe 242

## RESULT 3

G89873

hypothetical protein sspA [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: G89873

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: G89873

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-342 <KUR>

A:Cross-references: GB:BA000018; PID:gl3700850; PIDN:BA42146.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: sspA  
C:Superfamily: staphylococcal serine proteinase

Alignment Scores:  
Pred. No.: 0.512 Length: 342  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x G89873 (1-342)

Qy 1930 ACGGCGGTAACTCGGTAGCCCGTATTC 1959  
|||||

Db 233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 4

S21758

glutamic acid-specific endopeptidase - Staphylococcus aureus

C:Species: Staphylococcus aureus

C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Jun-2000

C:Accession: S21758

R:Yoshikawa, K.; Tsuzuki, H.; Fujiwara, T.; Nakamura, E.; Iwamoto, H.; Matsumoto, K.;

Biochim. Biophys. Acta 1121, 221-228, 1992

A:Title: Purification, characterization and gene cloning of a novel glutamic acid-spe

A:Reference number: S21758; MUID:92287954; PMID:1599945

A:Accession: S21758

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-357 <YOS>

A:Cross-references: GB:D00730; NID:g216970; PIDN:BAA00630.1; PID:g216971

C:Superfamily: staphylococcal serine proteinase

## Alignment Scores:

Pred. No.: 0.51 Length: 357  
Score: 10.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.40% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x S21758 (1-357)

Qy 1930 ACGGCGGTAACTCGGTAGCCCGTATTC 1959  
|||||

Db 233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 5

F75381

probable two-component sensor - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C:Accession: F75381

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J

, M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: F75381

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-569 <WHI>

A:Cross-references: GB:AE001999; GB:AE000513; NID:96459316; PIDN:AAF11119.1; PID:9645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1556

A:Map position: 1

Alignment Scores:  
 Pred. No.: 0.484 Length: 569  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x F75381 (1-569)

QY 25 CTTCTCGGAGCAGCCCTGCTTGGTGGT 54

Db 53 LeuLeuGlyAlaAlaLeuLeuGlyAla 72

RESULT 6

D75483 hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
 C:Accession: D75483  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: D75483

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-266 <WHI>

A:Cross-references: GB:AE001928; GB:AE000513; NID:96458421; PIDN:APF10302.1; PID:9645843

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0714

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DR0714

Alignment Scores:  
 Pred. No.: 5.59 Length: 266  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.26% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x D75483 (1-266)

QY 25 CTTCTCGGAGCAGCCCTGCTTGGGT 51

Db 108 LeuLeuGlyAlaAlaLeuLeuGly 116

RESULT 7

AE1710

hypothetical protein homolog lin2224 [imported] - Listeria innocua (strain Clip11262)

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AE1710

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AE1710

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-452 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC97453.1; PID:g16414737; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin2224

Alignment Scores:  
 Pred. No.: 5.27 Length: 452  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.26% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x AE1710 (1-452)

QY 1555 GTAGAGCTTCCAGAGCGTAATAGCT 1581

Db 142 ValGluLeuSerLysSerValleAla 150

RESULT 8

AF2363

hypothetical protein all14462 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AF2363

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AF2363

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-483 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA876161.1; PID:g17133598; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all14462

Alignment Scores:  
 Pred. No.: 5.23 Length: 483  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.26% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x AF2363 (1-483)

QY 34 GCAGCCCTGCTGTTGGTCTTCAGGG 60

Db 353 AlaAlaLeuLeuGlyAlaSerGly 361

RESULT 9

T36771

probable integral membrane protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T36771

R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A

submitted to the EMBL Data Library, July 1999

A:Reference number: 221574

A:Accession: T36771

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-684 <SEE>

A:Cross-references: EMBL:AL096844; PIDN:CAB50875.1; GSPDB:GN00070; SCOEDB:SCI28.01

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCI28.01

Alignment Scores:  
 Pred. No.: 5.03 Length: 634  
 Score: 9.00 Matches: 9  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.26% Indels: 0  
DB: Gaps: 0

US-10-008-355-1 (1-2139) x T36771 (1-684)  
QY 28 CTCGAGCAGCCCTGCTGTGGGTGCT 54  
Db 442 LeuGlyAlaLeuLeuLeuGlyAla 450

RESULT 10  
A34076  
protein-tyrosine kinase (EC 2.7.1.112) receptor type eph 1 precursor - human  
N:Alternate names: receptor tyrosine kinase eph  
C:Species: Homo sapiens (man)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 04-Feb-2000  
C:Accession: A34076; S44280  
R:Hiral, H.; Maru, Y.; Hagihara, K.; Nishida, J.; Takaku, F.  
Science 238, 1717-1720, 1987  
A:Title: A novel putative tyrosine kinase receptor encoded by the eph gene.  
A:Reference number: A34076; MUID:88070650; PMID:2825356  
A:Accession: A34076  
A:Molecule type: mRNA  
A:Residues: 1-984 <HIR>  
A:Cross-references: GB:M8391; NID:g339716; PIDN:AAA36747.1; PID:g339717  
A:Note: the sequence in GenBank entry HUMTKR, release 111.0, has the codons GCG for 398-  
R:Tuzi, N.L.  
submitted to the EMBL Data Library, November 1993  
A:Description: An EGFR/eph chimeric receptor possesses ligand stimulated tyrosine kinase  
A:Reference number: S44280  
A:Accession: S44280  
A:Molecule type: mRNA  
A:Residues: 286-397, 'A', 399-580, 'ORDRATDVDRDKLWKPVDLQAYEDPAQALDF', 583, 625-984 <TUZ>  
A:Cross-references: EMBL:Z27409; NID:g482916; PIDN:CA81796.1; PID:g482917  
C:Genetics:  
A:Gene: GDB:EPH1; EPH; EPH7  
A:Cross-references: GDB:I19875; OMIM:179610  
A:Map position: 7q32-7q36  
C:Superfamily: protein-tyrosine kinase, receptor type eph; fibronectin type III repeat  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-984/Product: protein-tyrosine kinase receptor type eph 1 #status predicted <MAT>  
F:548-568/Domain: transmembrane #status predicted <TM>  
F:630-895/Domain: protein kinase homology <KIN>  
F:638-646/Region: protein kinase ATP-binding motif  
F:918-984/Domain: SAM homology <SAM>  
F:59,338,414,478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:  
Pred. No.: 4.83 Length: 984  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.26% Indels: 0  
DB: Gaps: 1

US-10-008-355-1 (1-2139) x A34076 (1-984)  
QY 25 CTCCTCGAGCAGCCCTGCTGTGGGT 51  
Db 556 LeuLeuGlyAlaLeuLeuGlyAla 564

RESULT 11  
B70476  
hypothetical protein aq\_2054 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
C:Accession: B70476  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666; PMID:9537320  
A:Accession: B70476

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1116 <AQF>  
A:Cross-references: GB:AE000770; NID:g2984274; PIDN:AA07805.1; PID:g2984280; GB:AE00  
C:Experimental source: strain VF5  
C:Genetics:  
A:Gene: aq\_2054

Alignment Scores:  
Pred. No.: 4.76 Length: 1116  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.26% Indels: 0  
DB: Gaps: 2

US-10-008-355-1 (1-2139) x B70476 (1-1116)  
QY 1405 GTAATCGCAGCAAAATTCAAAGCGAC 1431  
Db 329 ValIleAspLysPheLysGlyAsp 337

RESULT 12  
T36095  
hypothetical protein SCE134.13 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 04-Mar-2000  
C:Accession: T36095  
R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M  
submitted to the EMBL Data Library, April 1999  
A:Reference number: Z21596  
A:Accession: T36095  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-102 <SAU>  
A:Cross-references: EMBL:AL049661; PIDN:CAB1211.1; GSPDB:GN000070; SCOEDB:SCE134.13  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCE134.13  
C:Superfamily: Streptomyces coelicolor hypothetical protein SCE134.13

Alignment Scores:  
Pred. No.: 65.9 Length: 102  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: Gaps: 2

US-10-008-355-1 (1-2139) x T36095 (1-102)  
QY 31 GGAGCAGCCCTGCTGTGGGTGCT 54  
Db 79 GlyAlaLeuLeuLeuGlyAla 86

RESULT 13  
T04804  
hypothetical protein F10M23.150 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
C:Accession: T04804  
R:Bevan, M.; Lecharny, A.; Chedford, F.; Krivitzky, M.; Kreis, M.; Hoheisel, J.; Mewes  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15385  
A:Accession: T04804  
A:Molecule type: DNA  
A:Residues: 1-106 <BEV>  
A:Cross-references: EMBL:AL035440  
A:Experimental source: cultivar Columbia; BAC clone F10M23  
C:Genetics:  
A:Map position: 4  
A:Introns: 58/3  
A:Note: F10M23.150

## Alignment Scores:

Pred. No.: 65.7 Length: 106  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x T04804 (1-106)

QY 10 AAATTAAAGTATTCTCTCGGA 33

Db 71 LysLeuLysSerIleLeuLeuGly 78

## RESULT 14

JC4669

cocaine-and amphetamine regulated transcript protein - human

C:Species: Homo sapiens (man)

C:Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 21-Jul-2000

R:Accession: JC4669; G01648

R:Dougllass, J.; Daoud, S.

Gene 169, 241-245, 1996

A:Title: Characterization of the human cDNA and genomic DNA encoding CART: A cocaine-

A:Reference number: JC4669; MUID:96194810; PMID:8647455

A:Accession: JC4669

A:Molecule type: mRNA

A:Residues: 1-116 &lt;DOU&gt;

A:Cross-references: GB:U16826; NID:g609305; PIDN:AAB08010.1; PID:g609306

R:Dougllass, J.O.

submitted to the EMBL Data Library, January 1995

A:Reference number: G08029

A:Accession: G01648

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-116 &lt;DO2&gt;

A:Cross-references: EMBL:U20325; NID:g665578; PID:g665579

C:Comment: This protein plays a conserved functional role within the mammalian central r

stimulants.

C:Genetics:

A:Gene: CART

A:Map position: 5

A:Introns: 53/3; 81/3

## Alignment Scores:

Pred. No.: 65 Length: 116  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x JC4669 (1-116)

QY 25 CTTCTGGAGCAGCCCTGCTGTG 48

Db 11 LeuLeuGlyAlaAlaLeuLeu 18

## RESULT 15

C69495

conserved hypothetical protein Afl964 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999

C:Accession: C69495

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: C69495

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-162 &lt;KLE&gt;

A:Cross-references: GB:AE000967; GB:AE000782; NID:g2689290; PIDN:AAB89289.1; PID:g264

## Alignment Scores:

Pred. No.: 62.6 Length: 162  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.13% Indels: 0  
DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x C69495 (1-162)

QY 1401 CTTGAATATATCGGGAGCTTGTG 1378

Db 142 LeuGluTyrIleGlyGluLeuVal 149

Search completed: May 16, 2003, 13:16:20  
Job time : 92 secs



GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:55:27 ; Search time 165 Seconds  
(without alignments)  
5342.240 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 712

Sequence: 1 atgcaaatgaattaaaag.....aagagctgaagtgatctaa 2139

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xip  
-Q/cgn2.1/USPTO.spool/US10008355/runat\_16052003.110426.9734/app.query.fasta\_1.2311  
-DB=SPTREMBL\_21 -QFMT=fastan -SUFFIX=olin2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355 -CGN\_1\_1\_238 -runat\_16052003.110426.9734 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGOQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rv.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	1.8	716	16 Q9PC94	Q9pc94 xylella fas

2	10	1.4	316	2	Q9FBG1	Q9fbg1 staphylococ
3	10	1.4	342	16	Q99V45	Q99v45 staphylococ
4	10	1.4	357	2	Q04186	Q04186 staphylococ
5	10	1.4	569	16	Q9RU39	Q9ru39 deinococcus
6	9	1.3	217	2	Q9AJX0	Q9ajx0 staphylococ
7	9	1.3	266	16	Q9RWF5	Q9rwf5 deinococcus
8	9	1.3	452	16	Q92900	Q92900 listeria in
9	9	1.3	483	16	Q8YNU9	Q8ynu9 anabaena sp
10	9	1.3	684	16	Q9S282	Q9s282 streptomyce
11	8	1.1	92	11	Q91WB5	Q91wb5 mus musculu
12	8	1.1	102	16	Q9X868	Q9x868 streptomyce
13	8	1.1	106	10	Q9S221	Q9s221 arabidopsis
c 14	8	1.1	162	17	Q28315	Q28315 archaeoglob
15	8	1.1	177	17	Q8ZVA1	Q8zva1 pyrobaculum
16	8	1.1	178	16	Q9K3J5	Q9k3j5 streptomyce
17	8	1.1	189	17	Q27455	Q27455 methanobact
18	8	1.1	199	10	Q9SS66	Q9ss66 arabidopsis
c 19	8	1.1	211	5	Q9V4H5	Q9v4h5 drosophila
c 20	8	1.1	224	16	Q9PJ38	Q9pj38 campylobact
21	8	1.1	226	4	Q8WZ66	Q8wz66 homo sapien
c 22	8	1.1	241	17	Q29854	Q29854 archaeoglob
c 23	8	1.1	268	17	Q8TV91	Q8tv91 methanopyru
c 24	8	1.1	284	10	Q9862	Q9862 glycine max
25	8	1.1	288	16	Q9X246	Q9x246 thermotoga
c 26	8	1.1	312	4	Q9NYW1	Q9nyw1 homo sapien
c 27	8	1.1	332	10	Q40600	Q40600 oenothera b
c 28	8	1.1	344	2	Q93CA1	Q93ca1 bifidobacte
29	8	1.1	350	9	Q9ZXM6	Q9zxm6 bacterioph
30	8	1.1	356	2	Q33999	Q33999 chromatium
31	8	1.1	372	5	Q9NGT6	Q9ngt6 leishmania
32	8	1.1	380	2	Q9FDK6	Q9fdk6 zymomonas m
33	8	1.1	391	5	Q93230	Q93230 caenorhabdi
c 34	8	1.1	392	16	Q9KBD1	Q9kbd1 bacillus fl
c 35	8	1.1	395	2	Q9AL66	Q9al66 shigella fl
c 36	8	1.1	396	2	Q8VNO8	Q8vno8 escherichia
37	8	1.1	396	4	Q96RL8	Q96rl8 homo sapien
38	8	1.1	396	4	Q8WWV3	Q8wwv3 homo sapien
39	8	1.1	400	2	Q9VV03	Q9vv03 streptomyce
40	8	1.1	400	16	Q92T10	Q92t10 rhizobium m
c 41	8	1.1	414	16	Q9F2U9	Q9f2u9 streptomyce
c 42	8	1.1	419	16	Q8U7A5	Q8u7a5 agrobacteri
c 43	8	1.1	421	2	Q9ADX1	Q9edx1 escherichia
c 44	8	1.1	421	2	Q8VQN9	Q8vqn9 escherichia
c 45	8	1.1	421	2	Q8VLM4	Q8vlm4 escherichia

ALIGNMENTS

RESULT 1					
Q9PC94	ID	Q9PC94	PRELIMINARY:	PRT:	716 AA.
AC	Q9PC94:				
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Hypothetical protein Xf1887.				
GN	Xf1887.				
OS	Xylella fastidiosa.				
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;				
CC	Xylella.				
OX	NCBI_TaxID=2371;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=9A5C;				
RX	MEDLINE=20365717; PubMed=10910347;				
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,				
RA	Alvarenga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,				
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,				
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,				
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,				
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,				
RA	Facincani A.P., Ferreira V.C.A., Ferreira J.A., Ferro J.A.,				
RA	Fraga J.S., Franca S.C., Frohme M., Furlan L.R.,				

RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kurama E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";  
 RL Nature 406:151-159(2000).  
 DR EMBL: AE004008: AAF84693.1; -  
 DR MEROPS: S46.001; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 716 AA; 79375 MW; EAF086E2315BDFC CRC64;

Alignment Scores:  
 Pred. No.: 0.000331 Length: 716  
 Score: 13.00 Matches: 13  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.83% Indels: 0  
 DB: 16 Gaps: 0

US-10-008-355-1 (1-2139) x Q9PC94 (1-716)

QY 643 GACAACGTGATCGCGCGTCACACGGCGCACTTCAGC 681  
 |||||||  
 DB 212 ASPASnTripMetTriProArgHisThrGlyAspPheSer 224

RESULT 2

Q9FBG1 ID Q9FBG1 PRELIMINARY: PRT; 316 AA.  
 AC Q9FBG1;  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE Glutamyl endopeptidase.  
 GN PROM.  
 OS Staphylococcus warneri.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 OX NCBI\_TaxID=1292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kakikawa M.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.

RA Kodaira K.I.;  
 RT "Characterization of the gene encoding glutamyl endopeptidase of  
 Staphylococcus warneri M.";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ293885; CAC06168.1; -  
 DR MEROPS: S01.269; -  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR InterPro: IPR000126; Ser\_proteas\_V8.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00839; V8PROTEASE.  
 DR SMART: SM00020; Tryp\_SPC; 1.  
 DR PROSITE: PS02040; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00673; V8\_SER; 1.  
 KW Hydrolase; Serine protease.  
 SQ SEQUENCE 316 AA; 34296 MW; 4E997A5A111DDB40 CRC64;

Alignment Scores:  
 Pred. No.: 0.599 Length: 316  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x Q9FBG1 (1-316)

QY 1930 ACGGCGGTAACTCGGTAGCCCGCTATTC 1959  
 |||||||  
 DB 231 ThrGlyGlyAsnSerGlySerProValPhe 240

RESULT 3

Q99V45 ID Q99V45 PRELIMINARY: PRT; 342 AA.  
 AC Q99V45;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Serine protease, V8 protease, glutamyl endopeptidase.  
 GN SSPA OR SAV1048 OR SA0301.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus.  
 OX NCBI\_TaxID=158878, 158879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S aureus (strain Mu50), and S aureus (strain N315);  
 RX MEDLINE=213111952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hitakawa H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Hara Y., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus.";  
 RL Lancet 357:1225-1240(2001).  
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
 DR EMBL: AP003361; BAB57210.1; -  
 DR EMBL: AP003132; BAB42146.1; -  
 DR MEROPS: S01.269; -  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR InterPro: IPR000126; Ser\_proteas\_V8.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00839; V8PROTEASE.  
 DR SMART: SM00020; Tryp\_SPC; 1.  
 DR PROSITE: PS02040; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00672; V8\_HIS; 1.  
 DR PROSITE: PS00673; V8\_SER; 1.  
 KW Hydrolase; Protease; Serine protease; Complete proteome.  
 SQ SEQUENCE 342 AA; 36977 MW; 5AEF42DC501C4B24 CRC64;

Alignment Scores:  
 Pred. No.: 0.594 Length: 342  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 16 Gaps: 0

US-10-008-355-1 (1-2139) x Q99V45 (1-342)

QY 1930 ACGGCGGTAACTCGGTAGCCCGCTATTC 1959  
 |||||||  
 DB 233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 4



```

Q04186
ID Q04186 PRELIMINARY; PRT; 357 AA.
AC Q04186;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Glutamic acid specific protease prepropeptide (EC 3.4.21.19).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoshikawa K., Tsuzuki H., Fujiwara T., Nakamura E., Iwamoto H.,
RA Matsumoto K., Shin M., Yoshida N., Teraoka H.;
RT "Purification, characterization and gene cloning of a novel glutamic
RT acid-specific endopeptidase from staphylococcus aureus atcc 12600.";
RL Biochim. Biophys. Acta 1121:221-228(1991).
DR EMBL: D00730; BAA00630.1; -.
DR MEROPS; S01.269; -.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000126; Ser_proteas_V8.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR PROSITE; PS02440; TRYPsin_DOM; 1.
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
DR Hydrolase; Protease; Serine protease.
KW Hydrolase; Protease; Serine protease.
FT CHAIN 69 357
FT POTENTIAL.
SQ SEQUENCE 357 AA; 38651 MW; 58AA9A4E371E2577 CRC64;

Alignment Scores:
Pred. No.: 0.591 Length: 357
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x Q04186 (1-357)
QY 1930 ACGGGCGTAACCTCCGTCAGCCCGTATTC 1959
Db 233 ThrGlyGlyAsnSerGlySerProValPhe 242

RESULT 5
Q9RU39 PRELIMINARY; PRT; 569 AA.
AC Q9RU39;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Two-component sensor, putative.
GN D11556.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Yamatehvan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RC Science 286:1571-1577(1999).
CC -1- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.

DR EMBL: AE001999; AAF11119.1; -.
DR TIGR: DRI556; -.
DR InterPro: IPR003594; ATPbind_ATPase.
DR InterPro: IPR004358; Bact_sens_pr_C.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR004359; HIS_KIN_sig.
DR Pfam: PF02518; HATPase_C; 1.
DR PRINTS; PR00344; BCTRLSENSOR.
DR SMART; SM00065; GAF; 1.
DR SMART; SM00387; HATPase_C; 1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase;
KW Complete proteome.
SQ SEQUENCE 569 AA; 59182 MW; 1A6CA4F88D96A940 CRC64;

Alignment Scores:
Pred. No.: 0.559 Length: 569
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 16 Gaps: 0

US-10-008-355-1 (1-2139) x Q9RU39 (1-569)
QY 25 CTTCTCGGACGACCGCTGCTGGTGGTGCT 54
Db 63 LeuLeuGlyAlaAlaLeuLeuLeuGlyAla 72

RESULT 6
Q9AJX0 PRELIMINARY; PRT; 217 AA.
AC Q9AJX0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Extracellular serine proteinase precursor (Fragment).
GN ESP.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6746;
RA Dubin G.;
RT "Staphylococcus epidermidis extracellular serine proteinase.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ305145; CAC27157.1; -.
DR HSSP; P09331; IEXF.
DR MEROPS; S01.269; -.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000126; Ser_proteas_V8.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR PROSITE; PS02440; TRYPsin_DOM; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT NON_TER 1
FT SIGNAL <1 1 POTENTIAL.
FT CHAIN 2 217 EXTRACELLULAR SERINE PROTEINASE.
SQ SEQUENCE 217 AA; 23667 MW; FB9B886D453B8BB7 CRC64;

Alignment Scores:
Pred. No.: 7.39 Length: 217
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.26% Indels: 0
DB: 2 Gaps: 0

US-10-008-355-1 (1-2139) x Q9AJX0 (1-217)
QY 1933 GCGGTAACCTCCGTCAGCCCGTATTC 1959

```

```

|||||
Db 167 GlycylAsnSerGlySerProValPhe 175
RESULT 7
Q9RWF5 PRELIMINARY; PRT; 266 AA.
AC Q9RWF5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein DR0714.
GN Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamatheva J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE001928; AAF10302.1; -.
DR TIGR; DR0714; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 266 AA; 28225 MW; F7C0ED7F231ADD3C CRC64;

Alignment Scores:
Pred. No.: 7.21 Length: 266
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.26% Indels: 0
DB: 16 Gaps: 0

US-10-008-355-1 (1-2139) x Q9RWF5 (1-266)
QY 25 CTCTCGGAGCAGCCCTGCTGTGGGT 51
Db 108 LeuLeuGlyAlaAlaLeuLeuGly 116
RESULT 8
Q929Q0 PRELIMINARY; PRT; 452 AA.
AC Q929Q0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein lin2224.
GN Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-W., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,

```

```

RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596171; CAC97453.1; -.
DR Listlist; LIN02224; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 452 AA; 48806 MW; 0CC0B8B8BF765745A CRC64;

Alignment Scores:
Pred. No.: 6.78 Length: 452
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.26% Indels: 0
DB: 16 Gaps: 0

US-10-008-355-1 (1-2139) x Q929Q0 (1-452)
QY 1555 GTAGAGCTTTCCAGAGCGTAATAGCT 1581
Db 142 ValGluLeuSerLysservAlaAla 150
RESULT 9
Q8YNU9 PRELIMINARY; PRT; 483 AA.
AC Q8YNU9;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein All4462.
GN All4462.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003596; BAH76161.1; -.
DR InterPro; IPR004896; PUCG.
DR Pfam; PF03209; PUCG; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 483 AA; 51883 MW; 17F82B83C39E7808 CRC64;

Alignment Scores:
Pred. No.: 6.72 Length: 483
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.26% Indels: 0
DB: 16 Gaps: 0

US-10-008-355-1 (1-2139) x Q8YNU9 (1-483)
QY 34 GCAGCCCTGCTGTGGTGGCTTCAGG 60
Db 353 AlaAlaLeuLeuGlyAlaSerGly 361
RESULT 10
Q9S282 PRELIMINARY; PRT; 684 AA.
ID Q9S282;
AC Q9S282;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Putative integral membrane protein.  
GN SCO1807 OR SCI28.01.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K., Harris D.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL096844; CAB50875.1; -.  
DR InterPro; IPR002965; P-rich\_extensn.  
DR PRINTS; PR01582; KV33CHANNEL.  
DR PROSITE; PRO1217; PRICHEXTENSIN.  
SQ SEQUENCE 684 AA; 66534 MW; F9FC89B471B99752 CRC64;  
  
Alignment Scores:  
Pred. No.: 6.45 Length: 684  
Score: 9.00 Matches: 9  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.26% Indels: 0  
DB: 16 Gaps: 0  
  
US-10-008-355-1 (1-2139) x Q9S282 (1-684)  
QY 28 CTCGACGACCCCTGCTGTGGTGCT 54  
Db 442 LeuGlyAlaLeuLeuLeuLeuLeuGlyAla 450  
  
RESULT 11  
Q91WB5 PRELIMINARY; PRT; 92 AA.  
AC Q91WB5;  
DT 01-DEC-2001 (TReMBLrel. 19, Created)  
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE Hypothetical 10.1 kDa protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SALIVARY GLAND;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC016132; AAH16132.1; -.  
DR InterPro; IPR002412; FEL\_allergen.  
DR InterPro; IPR000329; uteroglobin.  
DR Pfam; PF01099; Uteroglobin; 1.  
DR ProDom; PD019935; FEL\_allergen; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 92 AA; 10080 MW; FIA7557E0F5568BB CRC64;  
  
Alignment Scores:  
Pred. No.: 96.5 Length: 92  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: 11 Gaps: 0  
  
US-10-008-355-1 (1-2139) x Q91WB5 (1-92)  
QY 25 CTCCTCGACGACCCCTGCTGTG 48  
Db 9 LeuLeuGlyAlaAlaLeuLeuLeu 16  
  
RESULT 12  
Q9X868 PRELIMINARY; PRT; 102 AA.  
AC Q9X868;  
DT 01-NOV-1999 (TReMBLrel. 12, Created)  
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Hypothetical protein SCO3512.  
GN SCO3512 OR SCE134.13.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL049661; CAB41211.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 102 AA; 10274 MW; 848AD9BBE5308398 CRC64;  
  
Alignment Scores:  
Pred. No.: 95.3 Length: 102  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: 16 Gaps: 0  
  
US-10-008-355-1 (1-2139) x Q9X868 (1-102)  
QY 31 GGACGACCCCTGCTGTGGTGCT 54  
Db 79 GlyAlaAlaLeuLeuLeuGlyAla 86

## RESULT 13

Q95Z21 AC Q95Z21 PRELIMINARY; PRT; 106 AA.  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Hypothetical 12.2 kDa protein.  
 GN F10M23.150 OR AT4G26810.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Lecharny A., Chefdor F., Krivitzky M., Kreis M.,  
 RA Holsel J., Mewes H.W., Mayer K.F.X., Schueller C.;  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Lecharny A., Chefdor F., Krivitzky M., Kreis M., Mewes H.W.,  
 RA Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL035440; CAB36527.1;  
 DR EMBL: AL161565; CAB79536.1;  
 DR InterPro: IPR003121; SWIB.  
 DR Pfam: PF02201; SWIB; 1.  
 DR SMART: SM00151; SWIB; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 106 AA; 12209 MW; 8058B0ECA90F4CE1 CRC64;

Alignment Scores:  
 Pred. No.: 94.9 Length: 106  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.12% Indels: 0  
 DB: 10 Gaps: 0

US-10-008-355-1 (1-2139) x Q95Z21 (1-106)

QY 10 AAATTAAAGTATCTCTCGGA 33  
 Db 71 LysLeuYsSerIleLeuGly 78

## RESULT 14

O28315 AC O28315 PRELIMINARY; PRT; 162 AA.  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein AF1964.  
 GN AF1964.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 OC Archaeoglobaceae; Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kervavage A.R., Graham D.E., Kyriakides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing Archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 DR EMBL: AE00967; AAB89289.1;  
 DR TIGR: AF1964;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 162 AA; 18024 MW; 46404F9C3EE74AA CRC64;

Alignment Scores:  
 Pred. No.: 90.3 Length: 162  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.13% Indels: 0  
 DB: 17 Gaps: 0

US-10-008-355-1 (1-2139) x O28315 (1-162)

QY 1401 CTGCAATATATCGGGAGAGCTTGTC 1378  
 Db 142 LeuGluTyrIleGlyGluLeuVal 149

## RESULT 15

O82VAL ID O82VAL PRELIMINARY; PRT; 177 AA.  
 AC O82VAL;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Transport protein part 1, authentic frameshift.  
 GN PAE2385A.  
 OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 OX NCBI\_TaxID=13773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 DR EMBL: AE009873; AAL64155.1;  
 KW Complete proteome.  
 SQ SEQUENCE 177 AA; 19519 MW; FD2BB9B026DEC30F CRC64;

Alignment Scores:  
 Pred. No.: 89.3 Length: 177  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.12% Indels: 0  
 DB: 17 Gaps: 0

US-10-008-355-1 (1-2139) x O82VAL (1-177)

QY 1039 GACGTGATAGTCGTAAAGCGTGCC 1062  
 Db 69 AspValIleGlyArgLysArgAla 76

Search completed: May 16, 2003, 13:13:33  
Job time : 177 secs

---



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 16, 2003, 12:39:16 ; Search time 36.5 Seconds  
(without alignments)  
4861.253 Million cell updates/sec

Title: US-10-008-355-1  
Perfect score: 712  
Sequence: 1 atgcaaatgaaattaaaag.....aagagctgaagtgtatctaa 2139

Scoring table:  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues  
Word size: 1

Total number of hits satisfying chosen parameters: 225614  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlpl  
-O=/cgn2.1/USPTO.spool/US10008355/runat\_16052003\_110425\_9723/app\_query.fasta\_1.2311  
-DB=SwissProt40 -QWTF=fastan -SUFFIX=olin2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355.ecgn\_1.1.47 -runat\_16052003\_110425\_9723 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAEXT=60 -XGAPOP=6 -XGAEXT=7  
-YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	1.4	336	1. STSP_STA00	P04188 staphylococ
2	9	1.3	976	1. EPAL_HUMAN	P21709 homo sapien
3	9	1.3	1116	1. YK54_AQUAE	O67838 aquifex aeo
4	8	1.1	116	1. CART_HUMAN	Q16568 homo sapien
5	8	1.1	129	1. CART_MOUSE	P56388 mus musculus
6	8	1.1	214	1. TPIS_HALNI	Q9HQS4 halobacteri
7	8	1.1	226	1. ATP6_RAT	P05504 rattus norv
8	8	1.1	256	1. YDHI_HSVSC	P22575 herpesvirus
9	8	1.1	279	1. YN39_METKA	P58851 methanopyru
10	8	1.1	349	1. YNJ6_CAEEL	P34550 caenorhabdi
11	8	1.1	382	1. METK_SCHPO	O60198 schizosacch
12	8	1.1	392	1. CHST1_PERRA	Q9ZU06 persea amer
13	8	1.1	393	1. METK_ASCIM	P50304 ascobolus i
14	8	1.1	537	1. MPPI_ECOLI	P77348 escherichia
15	8	1.1	570	1. FBP3_STRPU	P49013 strongyloce
16	8	1.1	586	1. RUBA_ARATH	P21238 arabidopsis
17	8	1.1	637	1. ILV2_BRANA	P14874 brassica na
18	8	1.1	757	1. MUS2_THEMA	Q9X105 thermotoga

19	8	1.1	876	1	AREA_EMENI	P17429 emericella
20	8	1.1	900	1	SYA_AERPE	Q9Y9X3 aeropyrum p
21	8	1.1	969	1	MNR2_YEAST	P35724 saccharomyc
22	7	1.0	32	1	Y16Q_BPT4	P39247 bacterioph
23	7	1.0	42	1	RLA2_WHEAT	P05390 triticum ae
24	7	1.0	68	1	RL29_STRPY	Q9A1W6 streptococ
25	7	1.0	94	1	GATC_CAWJE	Q9P1A5 campylobact
26	7	1.0	113	1	PSAK_CHLRE	P14225 chlamydomon
27	7	1.0	116	1	Y1B6_MYCLE	O69519 mycobacteri
28	7	1.0	119	1	BCPI_BRACM	P41507 brassica ca
29	7	1.0	123	1	RS13_MYCGE	P71031 mycoplasma
30	7	1.0	124	1	URE2_BACSU	P17035 bacillus su
31	7	1.0	127	1	Y55B_MYCPN	P75047 mycoplasma
32	7	1.0	129	1	CART_RAT	P49192 rattus norv
33	7	1.0	136	1	SCPP_APLSP	P09892 aplysia sp.
34	7	1.0	139	1	NUSA_HALMO	P15738 halobacteri
35	7	1.0	139	1	NUSA_HALN1	P15739 halobacteri
36	7	1.0	142	1	H2B2_LYTP1	P06146 lytechinus
37	7	1.0	142	1	U426_HSVMG	Q05104 marek's dls
38	7	1.0	143	1	NUSA_SULAC	P11523 sulfolobus
39	7	1.0	147	1	HEAA_CAUCR	P27342 caulobacter
40	7	1.0	167	1	LSPA_CHLTR	O84413 chlamydia t
41	7	1.0	167	1	NU6M_BRALA	O79423 brachiosto
42	7	1.0	169	1	YSP0_STRCO	Q92B93 streptomyc
43	7	1.0	173	1	Y0J3_PSEAE	O9HXU7 pseudomonas
44	7	1.0	177	1	Y189_STRGR	Q9KH22 streptomyc
45	7	1.0	177	1	YH21_AZOGH	Q43924 azotobacter

ALIGNMENTS

RESULT 1	STSP_STA00	STANDARD;	PRT;	336 AA.
ID	STSP_STA00	STANDARD;	PRT;	336 AA.
AC	P04188:			
DT	20-MAR-1987 (Rel. 04, Created)			
DT	01-APR-1988 (Rel. 07, Last sequence update)			
DE	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Glutamyl endopeptidase precursor (EC 3.4.21.19) (Staphylococcal serine			
DE	proteinase) (V8 protease) (Endoproteinase Glu-C).			
GN	SSPA			
OS	Staphylococcus aureus.			
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.			
OX	NCBI_TaxID=1280;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=V8;			
RX	MEDLINE=87316953; PubMed=3306605;			
RA	Carmona C., Gray G.L.;			
RT	"Nucleotide sequence of the serine protease gene of Staphylococcus			
RT	aureus, strain V8."			
RL	Nucleic Acids Res. 15:6757-6757(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20569178; PubMed=11119502;			
RA	Rice K., Peralta R., Bast D., de Azavedo J., McGavin M.J.;			
RT	"Description of staphylococcus serine protease (ssp) operon in			
RT	staphylococcus aureus and nonpolar inactivation of sspa-encoded serine			
RT	protease."			
RL	Infect. Immun. 69:159-169(2001).			
RN	[3]			
RP	SEQUENCE OF 69-280.			
RC	STRAIN=V8;			
RX	MEDLINE=78212487; PubMed=96922;			
RA	Drapeau G.R.;			
RT	"The primary structure of staphylococcal protease."			
RL	Can. J. Biochem. 56:534-544(1978).			
CC	-I- FUNCTION: PREFERENTIALLY CLEAVES PEPTIDE BONDS ON THE CARBOXYL-			
CC	TERMINAL SIDE OF ASPARTATE AND GLUTAMATE.			
CC	-I- CATALYTIC ACTIVITY: Preferential cleavage: Asp-I-Xaa, Glu-I-Xaa.			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.			
CC	-I- DATABASE: NAME=worthington enzyme manual;			
CC	WWW="http://www.worthington-biochem.com/manual/P/STAP.html".			

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 EMBL: Y00356; CAA68434.1; - - - - -  
 EMBL: AF309515; AAG45843.1; - - - - -  
 PIR: A26812; PRSASK.  
 MEROPS: S01.269; - - - - -  
 InterPro: IPR00126; Ser\_proteas\_V8.  
 InterPro: IPR001254; Ser\_protease\_Try.  
 Pfam: PF00089; trypsin\_1.  
 PRINTS: PR00839; V8PROTEASE.  
 PROSITE: PS00672; V8\_HIS; 1.  
 PROSITE: PS00673; V8\_SER; 1.  
 Hydrolase: Serine protease; Zymogen; Signal.  
 SIGNAL 1 29  
 PROPEP 30 68  
 CHAIN 69 336  
 ACT\_SITE 119 161  
 ACT\_SITE 161 161  
 ACT\_SITE 237 237  
 CONFLICT 109 109  
 CONFLICT 125 125  
 CONFLICT 145 145  
 CONFLICT 193 193  
 CONFLICT 229 229  
 CONFLICT 259 261  
 CONFLICT 268 270  
 ENV -> NEVN (IN REF. 3).  
 ENV -> NEVN (IN REF. 3).  
 ENV -> NEVN (IN REF. 3).  
 SQ SEQUENCE 336 AA; 36326 MW; 8B138D0C7996AA3E CRC64;

Alignment Scores:  
 Pred. No.: 0.228 Length: 336  
 Score: 10.00 Matches: 10  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.40% Indels: 0  
 DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x STSP\_STA00 (1-336)

Oy 1930 ACGGCGGTAACTCCGGTAGCCCGTATTC 1959

Db 233 ThrGlyGlyAsnSerGlySerProvalPhe 242

RESULT 2

EPAL\_HUMAN STANDARD; PRT; 976 AA.  
 AC P21709; Q15405;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ephrin type-A receptor 1 precursor (EC 2.7.1.112) (Tyrosine-protein  
 kinase receptor EPH)  
 GN EPAL OR EPHT1 OR EPHT OR EPH.  
 OS Homo.sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8070650; PubMed=2825356;  
 RA Hirai H., Maru Y., Hagiwara K., Nishida J., Takaku F.;  
 RT "A novel putative tyrosine kinase receptor encoded by the eph gene.";  
 RL Science 238:1717-1720(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92299440; PubMed=10369740;  
 RA Owsalimpur D., Kelley M.J.;

RT \*Genomic structure of the EPAL receptor tyrosine kinase gene.\*;  
 RN Mol. Cell. Probes 13:169-173(1999).  
 RP SEQUENCE OF 286-976 FROM N.A.  
 RC TISSUE-Placenta;  
 RA Tuzi N.L.;  
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: RECEPTOR FOR MEMBERS OF THE EPHRIN-A FAMILY. BINDS WITH  
 A LOW AFFINITY TO EPHRIN-A1.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 tyrosine phosphate.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- TISSUE SPECIFICITY: OVEREXPRESSED IN SEVERAL CARCINOMAS.  
 CC -!- SIMILARITY: CONTAINS 1 SAM DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN  
 RECEPTOR SUBFAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 EMBL: M18391; AAA36747.1; ALT\_SEQ.  
 EMBL: AF101171; AAD43440.1; - - - - -  
 EMBL: AF101165; AAD43440.1; JOINED.  
 EMBL: AF101166; AAD43440.1; JOINED.  
 EMBL: AF101167; AAD43440.1; JOINED.  
 EMBL: AF101168; AAD43440.1; JOINED.  
 EMBL: AF101169; AAD43440.1; JOINED.  
 EMBL: AF101170; AAD43440.1; JOINED.  
 EMBL: Z27409; CAA81796.1; - - - - -  
 PIR: A34076; A34076.  
 DR HSP; P00523; 2PTK.  
 DR Genew; HGNC:3385; EPHA1.  
 DR MIM; 179610; - - - - -  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR001090; Ephrin\_receptor.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR001660; SAM.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR001426; Ykase\_receptorV.  
 DR Pfam; PF00041; fn3; 2.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00536; SAM; 1.  
 DR Pfam; PF01404; EPH\_lbd; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR PRODOM; PD000001; Euk\_pkinase; 1.  
 DR PRODOM; PD001495; Ephrin\_receptor; 1.  
 DR SMART; SM00060; FN3; 2.  
 DR SMART; SM00454; SAM; 1.  
 DR SMART; SM00219; TYRK; 1.  
 DR PROSITE; PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00790; RECEPTOR\_TYR\_KIN\_V\_1; 1.  
 DR PROSITE; PS00791; RECEPTOR\_TYR\_KIN\_V\_2; 1.  
 DR PROSITE; PS0105; SAM\_DOMAIN; 1.  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Receptor; Transmembrane; Glycoprotein; Signal; Repeat.  
 FT SIGNAL 1 23  
 FT CHAIN 24 976  
 FT DOMAIN 24 547  
 FT DOMAIN 548 568  
 FT TRANSMEM 569 976  
 FT DOMAIN 191 329  
 FT DOMAIN 330 442  
 FT DOMAIN 443 537  
 FT DOMAIN 537 537





ventrolateral part of the arcuate nucleus, in the external zone of the median eminence, and also found in terminals in the periventricular part of the paraventricular nucleus.

-!- INDUCTION: By leptin.

-!- DISEASE: A defect in CART is associated with reduced resting energy expenditure and cosegregates with obesity phenotype.

-!- SIMILARITY: BELONGS TO THE CART FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; U16826; AAB08010.1; -  
 DR EMBL; U20325; AAB08011.1; -  
 DR PDB; 1HY9; 29-AUG-01.  
 DR MIM; 602606; -  
 KW Neurotensin; Neurotransmitter; Cleavage on pair of basic residues;  
 KW Signal; Disease mutation; Polymorphism; Obesity; 3D-structure.  
 FT SIGNAL 1 27  
 FT CHAIN 28 116  
 FT COCAINE- AND AMPHETAMINE-REGULATED  
 FT TRANSCRIPT PROTEIN.  
 FT PEPTIDE 28 66  
 FT PEPTIDE 69 116  
 FT DISULFID 82 100  
 FT DISULFID 88 108  
 FT DISULFID 102 115  
 FT VARIANT 61 61  
 FT VARIANT 66 66  
 FT L -> F (IN AN OBESE PATIENT).  
 FT /FTID=VAR\_012199.  
 FT S -> T.  
 FT /FTID=VAR\_012200.  
 SQ SEQUENCE 116 AA; 12829 MW; FC396CA2C032AA83 CRC64;  
 /FTID=VAR\_012200.

Alignment Scores:  
 Pred. No.: 30.1 Length: 116  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.12% Indels: 0  
 DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x CART\_HUMAN (1-116)

Oy 25 CTTCTCGGAGCAGCCCTGCTGTG 48  
 |||||||||||||||||||||||||||||  
 Db 11 LeuLeuGlyAlaAlaLeuLeuLeu 18

RESULT 5  
 CART\_MOUSE  
 ID CART\_MOUSE STANDARD; PRT; 129 AA.  
 AC P56388; Q9QX28;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cocaine- and amphetamine-regulated transcript protein precursor  
 DE [Contains: CART(1-52); CART(55-102)].  
 GN CART.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC STRAIN=129/SVJ;  
 RX MEDLINE=20080750; PubMed=10612705;  
 RA Adams L.D., Gong W., Vechia S.D., Hunter R.G., Kuhar M.J.;  
 RT "CART: from gene to function";  
 RL Brain Res. 848:137-140(1999).  
 CC -!- FUNCTION: Satiety factor closely associated with the actions of  
 leptin and neurotensin; this anorectic peptide inhibits both

normal and starvation-induced feeding and completely blocks the feeding response induced by neurotensin Y and regulated by leptin in the hypothalamus (By similarity).

-!- SUBCELLULAR LOCATION: Secreted (Potential).

-!- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a short form; are produced by alternative splicing.

-!- SIMILARITY: BELONGS TO THE CART FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AF148071; AAF24168.1; -  
 DR HSSP; Q16568; 1HY9.  
 DR MGD; MGI:1351330; Cart.  
 KW Neurotensin; Neurotransmitter; Cleavage on pair of basic residues;  
 KW Alternative splicing; Signal.  
 FT SIGNAL 1 27  
 FT CHAIN 28 129  
 FT COCAINE- AND AMPHETAMINE-REGULATED  
 FT TRANSCRIPT PROTEIN.  
 FT PEPTIDE 28 79  
 FT PEPTIDE 82 129  
 FT PEPTIDE 89 129  
 FT DISULFID 95 113  
 FT DISULFID 101 121  
 FT DISULFID 115 128  
 FT VARSPLC 54 66  
 FT SEQUENCE 129 AA; 14285 MW; E4CA6CE70BEE6DF2 CRC64;  
 /FTID=VAR\_012199.

Alignment Scores:  
 Pred. No.: 29.7 Length: 129  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.12% Indels: 0  
 DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x CART\_MOUSE (1-129)

Oy 25 CTTCTCGGAGCAGCCCTGCTGTG 48  
 |||||||||||||||||||||||||||||  
 Db 11 LeuLeuGlyAlaAlaLeuLeuLeu 18

RESULT 6  
 TPIS\_HALN1  
 ID TPIS\_HALN1 STANDARD; PRT; 214 AA.  
 AC Q9HQS4;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Triosephosphate isomerase (EC 5.3.1.1) (TIM).  
 GN TPIA OR VNG1027G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 ON NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Eberhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).

CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone  
phosphate.  
CC -!- PATHWAY: Plays an important role in several metabolic pathways.  
CC -!- SUBUNIT: Homotetramer (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE005037; AAG19439.1; -  
DR InterPro; IPR003009; FMN\_enzyme.  
DR InterPro; IPR002173; pfkB.  
DR InterPro; IPR000652; Triophos\_ismrse.  
DR InterPro; PD001005; Triophos\_ismrse; 1.  
DR TIGRfams; TIGR00419; tm; 1.  
DR PROSITE; PS00171; TIM; FALSE\_NEG.  
KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;  
KW Pentose shunt; Complete proteome.  
FT ACT\_SITE 85 85  
SQ SEQUENCE 214 AA; 20928 MW; B6DD20BDD85D6A4 CRC64;  
  
Alignment Scores:  
Pred. No.: 27.7 Length: 214  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: 1 Gaps: 0  
  
US-10-008-355-1 (1-2139) x TPIS\_HALM1 (1-214)  
QY 52 GCTTCAGGGGTGACCAAGCCGAC 75  
Db 194 AlaSerGlyValAlaLysAlaAsp 201  
RESULT 7  
ID ATP6\_RAT STANDARD; PRT; 226 AA.  
AC P05504; O972E9;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE ATP synthase A chain (EC 3.6.3.14) (Protein 6).  
GN MTATP6 OR ATP6.  
OS Rattus norvegicus (Rat).  
OG Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Wistar;  
RX MEDLINE=89362487; PubMed=2504926;  
RA Gadaleta G., Pepe G., de Candia G., Quagliariello C., Sbisa E.,  
RA Saccocc C.;  
RT "The complete nucleotide sequence of the Rattus norvegicus  
mitochondrial genome: cryptic signals revealed by comparative  
analysis between vertebrates.";  
RL J. Mol. Evol. 28:497-516(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RA Grosskopf R., Feldmann H.;  
RT "Analysis of a DNA segment from rat liver mitochondria containing the  
genes for the cytochrome oxidase subunits I, II, II, ATPase subunit  
6, and several tRNA genes.";  
RL Curr. Genet. 4:151-158(1981).  
RN [3]

RP SEQUENCE FROM N.A., AND VARIANT ASN-101.  
RC STRAIN=BHE/CDB, and Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=96118462; PubMed=8529844;  
RA Mathews C.E., McGraw R.A., Berdanier C.D.;  
RT "A point mutation in the mitochondrial DNA of diabetes-prone BHE/cdb  
rats.";  
RL FASEB J. 9:1638-1642(1995).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.55 ANGSTROMS) OF 29-41.  
RX MEDLINE=21109741; PubMed=11163232;  
RA Speir J.A., Stevens J., Joly E., Butcher G.W., Wilson I.A.;  
RT "Two different, highly exposed, bulged structures for an unusually  
long peptide bound to rat MHC class I RT1-A(a).";  
RL Immunity 14:81-92(2001).  
CC -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A  
DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +  
H(+)(Out).  
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC  
CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)  
HAS THREE MAIN SUBUNITS: A, B AND C.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; J01435; AAD15019.1; -  
DR EMBL; X14848; CAA32959.1; -  
DR EMBL; AF115770; AAD22965.1; -  
DR PIR; S04752; S04752.  
DR PDB; 1ED3; 21-JUN-00.  
DR InterPro; IPR000568; ATPsynth\_Asub.  
DR Pfam; PF00119; ATP-synt\_A; 1.  
DR PRINTS; PR00123; ATPASEA.  
DR TIGRfams; TIGR01131; ATP\_synth\_6\_or\_A; 1.  
DR PROSITE; PS00449; ATPASE\_A; 1.  
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane;  
KW 3D-structure.  
FT VARIANT 101 101 D -> N.  
FT CONFLICT 7 7 A -> P (IN REF. 2).  
FT CONFLICT 129 129 S -> L (IN REF. 2).  
FT CONFLICT 205 205 A -> V (IN REF. 2).  
SQ SEQUENCE 226 AA; 25050 MW; 6074E2CCAC2B586E CRC64;  
  
Alignment Scores:  
Pred. No.: 27.4 Length: 226  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.13% Indels: 0  
DB: 1 Gaps: 0  
  
US-10-008-355-1 (1-2139) x ATP6\_RAT (1-226)  
QY 1517 TTGTCGATGGACTTGACGATGCA 1494  
Db 98 LeuSerMetAspLeuSerMetAla 105  
RESULT 8  
ID YDHL\_HSVSC  
AC P22575;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 28.7 kda protein in DHFR 3' region (ORF1).

```
OS Herpesvirus saimiri (subgroup C / strain 488).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10384;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90266466; PubMed=2161148;
RA Blesinger B., Trimble J.J., Desrosiers R.C., Fleckenstein B.;
RT "The divergence between two oncogenic Herpesvirus saimiri strains in
RL a genomic region related to the transforming phenotype.";
RL Virology 176:505-514(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55264; AAA72928.1; -.
DR PIR; A34770; A34770.
KW Hypothetical protein.
SQ SEQUENCE 256 AA; 28662 MW; E04BA49D27A59D3A CRC64;
Alignment Scores:
Pred. No.: 27 Length: 256
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.12% Indels: 0
DB: 1 Gaps: 0
US-10-008-355-1 (1-2139) x YDHL_HSVSC (1-256)
QY 548 ACAAGCAATCTCTCTCTCTCTCT 571
Db 33 ThrThrAsnThrSerSerSer 40
RESULT 9.
Y939_METKA
ID Y939_METKA STANDARD; PRT; 279 AA.
AC P38851;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wooldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: CONTAINS 1 EH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29095; CAA82351.1; -.
DR PIR; S40718; S40718.
DR WormPep; R10E11.6; CE00308.
DR InterPro; IPR000261; EPS15_repeat.
DR PROSITE; PS50031; EH; 1.
KW Hypothetical protein.
FT DOMAIN 43 148 EH
SQ SEQUENCE 349 AA; 37984 MW; 862F7D9D3E3E2E2F CRC64;
Alignment Scores:
Pred. No.: 25.8 Length: 349
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
US-10-008-355-1 (1-2139) x Y939_METKA (1-279)
QY 301 GTGATCACCACATCTCTCGCGCAT 324
Db 224 ValAspHisAspTyrLeuArgAsp 231
RESULT 10
Y939_CAEEL
ID Y939_CAEEL STANDARD; PRT; 349 AA.
AC P34550;
RA 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 38.0 kDa protein R10E11.6 in chromosome III.
GN R10E11.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wooldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: CONTAINS 1 EH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z29095; CAA82351.1; -.
DR PIR; S40718; S40718.
DR WormPep; R10E11.6; CE00308.
DR InterPro; IPR000261; EPS15_repeat.
DR PROSITE; PS50031; EH; 1.
KW Hypothetical protein.
FT DOMAIN 43 148 EH
SQ SEQUENCE 349 AA; 37984 MW; 862F7D9D3E3E2E2F CRC64;
Alignment Scores:
Pred. No.: 25.8 Length: 349
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
```

```
Query Match: 1.12% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x YNJ6_CAEEL (1-349)

OY 545 CCAACACGATATCTCCATCG 568
DB 150 ProThrThrAsnThrSerSer 157

RESULT 11
ID METK_SCHPO STANDARD; PRT; 382 AA.
AC 060198;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
DE adenosyltransferase) (AdoMet synthetase).
GN SAM1 OR SPBC14F5.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Hilti N.;
RT "sam1, a new gene coding for S-adenosylmethionine synthetase in S.
RT pombe.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Borzym K., Langer I., Beck A., Holzer E., Moestl D., Hilbert H.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Pohl T.M.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga K.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -|- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
CC METHIONINE AND ATP (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
CC diphosphate + S-adenosyl-L-methionine.
CC -|- PATHWAY: Activated methyl cycle.
CC -|- SIMILARITY: BELONGS TO THE ADO MET SYNTHETASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
-----
CC EMBL; AD001672; AAC98143.1;
CC InterPro; IPR001099; N-C-synthase.
CC Pfam; PF00195; Chal_stil_synth; 1.
-----
CC or send an email to license@isb-sib.ch.
-----
CC EMBL; AJ001705; CAA04941.1;
CC EMBL; AL023780; CAA19323.1;
CC HSSP; P04384; 1MXB.
CC InterPro; IPR002133; S-AdoMet_synt.
CC Pfam; PF00438; S-AdoMet_synt; 1.
CC Pfam; PF02772; S-AdoMet_syntD2; 1.
CC Pfam; PF02773; S-AdoMet_syntD3; 1.
CC TIGRFAMs; TIGR01034; metK; 1.
CC PROSITE; PS00376; ADO MET SYNTHETASE_1; 1.
CC PROSITE; PS00377; ADO MET SYNTHETASE_2; 1.
KW Transferase; One-carbon metabolism; ATP-binding.
FT NP_BIND 118 123 ATP (POTENTIAL).
FT BINDING 146 146 ATP (POTENTIAL).
SQ SEQUENCE 382 AA; 41831 MW; 9970A9D1195C5738 CRC64;

Alignment Scores: 25.5 Length: 382
Pred. No.: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.12% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x METK_SCHPO (1-382)

OY 1755 CGTGCGCTGTACAACTATCATAC 1778
DB 300 ArgCysLeuValGlnLeuSerTyr 307

RESULT 12
CHSY_PERAEE STANDARD; PRT; 392 AA.
AC Q9ZU06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chalcone synthase (EC 2.3.1.74) (Naregenin-chalcone synthase).
GN CHS.
OS Persea americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
RN [1]
RP SEQUENCE FROM N.A.
RA Ardi R., Kobler I., Jacoby B., Keen N.T., Prusky D.;
RT "Involvement of epicatechin biosynthesis in the activation of the
RT mechanism of resistance of avocado fruits to Colletotrichum
RT gloeosporioides.";
RL Physiol. Mol. Plant Pathol. 53:269-286(1998).
CC -|- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE).
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -|- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA +
CC naringenin-chalcone + 3 CO(2).
CC -|- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -|- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
-----
CC EMBL; AD001672; AAC98143.1;
CC InterPro; IPR001099; N-C-synthase.
CC Pfam; PF00195; Chal_stil_synth; 1.
```

DR Pfam: PF02797; Chal\_stil\_synthC; 1.  
 DR PROSITE; PD000453; N-C\_synthase; 1.  
 DR PROSITE; PS00441; CHALCONE\_SYNTH; 1.  
 KW Flavonoid biosynthesis: Transferase; Acyltransferase.  
 FT ACT\_SITE 165 165 BY SIMILARITY.  
 SQ SEQUENCE 392 AA; 42480 MW; B08D286FB80BB42E CRC64;

## Alignment Scores:

Pred. No.: 25.4 Length: 392  
 Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.12% Indels: 0  
 DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x CHSV\_PRAE (1-392)

QY 1223 GTTTGTCACAGTTTGCCACGCAT 1246  
 |||||  
 Db 385 ValLeuHisSerLeuProThrHis 392

## RESULT 13

METK\_ASCIM STANDARD; PRT; 393 AA.  
 AC P50304;

DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine  
 adenosyltransferase) (Adomet synthetase).  
 OS Ascobolus immersus.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;  
 OC Pezizales; Ascobolaceae; Ascobolus.  
 OX NCBI\_TaxID=5191;  
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RN42;

RA MEDLINE=96200878; PubMed=8621082;

RA Mautino M.R., Goyon C., Rosa A.L.;

RT "Cloning and sequence of the Ascobolus immersus S-adenosyl-L-

methionine synthetase-encoding gene.";

RL Gene 170:155-156(1996).

CC -1- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM

CC METHIONINE AND ATP.

CC -1- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +

CC diphosphate + S-adenosyl-L-methionine.

CC -1- PATHWAY: Activated methyl cycle.

CC -1- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U21548; AAB03805.1; -

DR HSP; P04384; IMXB.

DR InterPro: IPR002133; S-AdoMet\_synth.

DR Pfam; PF00438; S-AdoMet\_synth; 1.

DR Pfam; PF02772; S-AdoMet\_synthD2; 1.

DR Pfam; PF02773; S-AdoMet\_synthD3; 1.

DR TIGRFAMs; TIGR01034; metK; 1.

DR PROSITE; PS00376; ADOMET\_SYNTHETASE\_1; 1.

DR PROSITE; PS00377; ADOMET\_SYNTHETASE\_2; 1.

KW Transferase: One-carbon metabolism: ATP-binding.

FT NP\_BIND 127 132 ATP (POTENTIAL).

FT BINDING 155 155 ATP (POTENTIAL).

SQ SEQUENCE 393 AA; 43010 MW; 88FB2F2F14B751C3 CRC64;

## Alignment Scores:

Pred. No.: 25.4 Length: 393

Score: 8.00 Matches: 8  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 1.12% Indels: 0  
 DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x METK\_ASCIM (1-393)

QY 1755 CGTGCTCGTCACTATCATCAT 1778  
 |||||

Db 309 ArgCysLeuValGlnLeuSerTyr 316

## RESULT 14

MPPA\_ECOLI

ID MPPA\_ECOLI STANDARD; PRT; 537 AA.

AC P77348;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Periplasmic murein peptide-binding protein precursor.

GN MPPA OR B1329.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-32.

RC STRAIN=K12 / AT980;

RA MEDLINE=98155149; PubMed=9495761;

RA Park J.T., Raychaudhuri D., Li H., Normark S., Mengin-Lecreux D.;

RT "MppA, a periplasmic binding protein essential for import of the

RT bacterial cell wall peptide L-alanyl-gamma-D-glutamyl-meso-

RT diaminopimelate.";

RL J. Bacteriol. 180:1215-1223(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RA MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RA MEDLINE=97251357; PubMed=9097039;

RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,

RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,

RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,

RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,

RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,

RA Takenoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

RT corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).

CC -1- FUNCTION: ESSENTIAL FOR THE UPTAKE OF THE MUREIN PEPTIDE L-ALANYL-

CC GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE. ALSO TRANSPORTS SOME ALPHA-

CC LINKED PEPTIDES SUCH AS PRO-PHE-LYS WITH LOW AFFINITY. THE

CC TRANSPORT IS EFFECTED BY THE OLIGOPEPTIDE PERMEASE SYSTEM.

CC -1- SUBCELLULAR LOCATION: Periplasmic.

CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING

CC PROTEIN FAMILY 5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U88242; AAC38216.1; .  
DR EMBL; AE000231; AAC74411.1; ALT\_INIT.  
DR EMBL; D90772; BAA14932.1; ALT\_INIT.  
DR EMBL; D90771; BAA14922.1; ALT\_INIT.  
DR HSP; P06202; IJEV.  
DR SWISS-2DPAGE; P77348; COLI.  
DR EcoGene; EG13376; mppA.  
DR InterPro; IPR000914; SBP\_bac\_5.  
DR Pfam; PF00496; SBP\_bac\_5; 1.  
DR PROSITE; PS01040; SBP\_BACTERIAL\_5; 1.  
KW Peptide transport; Periplasmic; Signal; Complete proteome.  
FT SIGNAL 1 22 PERIPLASMTIC MUREIN PEPTIDE-BINDING  
FT CHAIN 23 537 PROTEIN.  
SQ SEQUENCE 537 AA; 59900 MW; C6A17656836DC3AC CRC64;  
  
Alignment Scores:  
Pred. No.: 24.4 Length: 537  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.12% Indels: 0  
DB: 1 Gaps: 0  
  
US-10-008-355-1 (1-2139) x MPPA\_ECOLI (1-537)  
QY 1892 GTCAGCTCCATATCGCTTCTCAT 1915  
Db 14 ValSerSerLeuSerLeuSerTyr 21  
  
RESULT 15  
FBP3\_STRPU STANDARD; PRT; 570 AA.  
AC P49013;  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Fibropellin C precursor (Epidermal growth factor-related protein 3)  
DE (EGF III) (Fibropellin III).  
GN EGF3.  
OS Strongylocentrotus purpuratus (Purple sea urchin).  
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;  
OC Strongylocentrotus.  
OX NCBI\_TaxID=7668;  
RN [1]  
RS SEQUENCE FROM N.A.  
RC TISSUE-Gastrula;  
RX MEDLINE=93273088; PubMed=8500658;  
RA Blagrove B.W., Raff R.A.;  
RT "The SpEGF III gene encodes a member of the fibropellins: EGF repeat-  
RT containing proteins that form the apical lamina of the sea urchin  
RT embryo."  
RL Dev. Biol. 157:526-538(1993).  
CC -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR  
CC MATRIX.  
CC -!- SUBCELLULAR LOCATION: Extracellular.  
CC -!- DEVELOPMENTAL STAGE: LOW LEVELS IN UNFERTILIZED EGGS AND DURING  
CC EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN LATE  
CC MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS MAINTAINED  
CC THROUGH SUBSEQUENT STAGES.  
CC -!- MISCELLANEOUS: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.  
CC -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.  
CC -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR  
CC TO AVIDIN/SPREPTAVIDIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
-----  
CC EMBL; L07045; AAA30045.1; .  
CC HSP; P00740; IEDM.  
DR InterPro; IPR000152; Asx\_hydroxyl.  
DR InterPro; IPR000085; Avidin.  
DR InterPro; IPR000859; CUB\_domain.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR000742; EGF-2.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR001438; EGF\_II.  
DR Pfam; PF00008; EGF; 8.  
DR Pfam; PF00431; CUB; 1.  
DR Pfam; PF01382; Avidin; 1.  
DR PRINTS; PR00010; EGF\_BLOOD.  
DR SMART; SM00042; CUB; 1.  
DR SMART; SM00179; EGF\_CA; 7.  
DR SMART; SM00001; EGF-Like; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 8.  
DR PROSITE; PS00022; EGF\_1; 8.  
DR PROSITE; PS00577; AVIDIN; 1.  
DR PROSITE; PS01180; CUB; 1.  
DR PROSITE; PS01186; EGF-2; 7.  
DR PROSITE; PS01187; EGF\_CA; 6.  
KW Biotin; EGF-like domain; Repeat; Signal; Glycoprotein;  
KW Calcium-binding.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 570 FIBROPELLIN C.  
FT DOMAIN 18 55 EGF-LIKE 1.  
FT DOMAIN 62 175 CUB.  
FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 366 402 EGF-LIKE 7.  
FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 442 570 AVIDIN-LIKE.  
FT DISULFID 23 34 BY SIMILARITY.  
FT DISULFID 28 43 BY SIMILARITY.  
FT DISULFID 45 54 BY SIMILARITY.  
FT DISULFID 180 191 BY SIMILARITY.  
FT DISULFID 185 200 BY SIMILARITY.  
FT DISULFID 202 211 BY SIMILARITY.  
FT DISULFID 218 229 BY SIMILARITY.  
FT DISULFID 223 238 BY SIMILARITY.  
FT DISULFID 240 249 BY SIMILARITY.  
FT DISULFID 256 267 BY SIMILARITY.  
FT DISULFID 261 276 BY SIMILARITY.  
FT DISULFID 278 287 BY SIMILARITY.  
FT DISULFID 294 305 BY SIMILARITY.  
FT DISULFID 299 314 BY SIMILARITY.  
FT DISULFID 316 325 BY SIMILARITY.  
FT DISULFID 332 343 BY SIMILARITY.  
FT DISULFID 337 352 BY SIMILARITY.  
FT DISULFID 354 363 BY SIMILARITY.  
FT DISULFID 370 381 BY SIMILARITY.  
FT DISULFID 375 390 BY SIMILARITY.  
FT DISULFID 392 401 BY SIMILARITY.  
FT DISULFID 408 419 BY SIMILARITY.  
FT DISULFID 413 428 BY SIMILARITY.  
FT DISULFID 430 439 BY SIMILARITY.  
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 570 AA; 61116 MW; BE65E3E1C05E6EE CRC64;  
  
Alignment Scores:  
Pred. No.: 24.2 Length: 570  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.13% Indels: 0

DB: 1 Gaps: 0

US-10-008-355-1 (1-2139) x FBP3\_STRPU (1-570)

QY 1381 TGTGGGAGGGATACGCGGGGTA 1358

Db 430 CysArgGlnGlyTyrAlaGlyval 437

Search completed: May 16, 2003, 13:07:56  
Job time : 48.5 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 01:38:11 ; Search time 3816 Seconds  
(without alignments)  
16313.129 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaaatgaaatgaaag.....aagagctgaagtgtgatctaa 2139

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_cm.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_frod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	129.4	6.0	10689	1	AE004008	AE004008 Xylella f
C 2	61.6	2.9	10811	1	AE011732	AE011732 Xanthomon
C 3	43.2	2.0	125020	9	AF429315	AF429315 Homo sapi
C 4	42.6	2.0	820	14	AF403408	AF403408 Golden sh
C 5	42.2	2.0	1461	3	AY051790	AY051790 Drosophil
C 6	42.2	2.0	76748	2	AC014256	AC014256 Drosophil
C 7	42.2	2.0	148847	3	AC008234	AC008234 Drosophil
C 8	42.2	2.0	162593	3	AC008356	AC008356 Drosophil
C 9	42.2	2.0	224400	3	AE003682	AE003682 Drosophil
C 10	41.6	1.9	720	8	CNSOLDBA	AL116606 Botrytis
C 11	41	1.9	1700	1	RHMTTRBC	AL5810 R.melloti
C 12	41	1.9	329100	1	SME591787	AL591787 Sinorhizo
C 13	40.4	1.9	125020	9	AF429315	AF429315 Homo sapi
C 14	40	1.9	4314	5	AF051784	AF051784 Xenopus l
C 15	40	1.9	202306	2	AC121850	AC121850 Mus muscu
C 16	39.4	1.8	3060	8	EMEFLLBA	L24395 Emericella
C 17	38.8	1.8	348450	1	MLEPRTN4	AL583920 Mycobacte
C 18	38.6	1.8	13155	1	PDSOX	X79242 Paracoccus
C 19	38.6	1.8	155962	9	AC099055	AC099055 Homo sapi
C 20	38.4	1.8	100986	8	AC105363	AC105363 Oryza sat
C 21	38.4	1.8	126637	2	AC118980	AC118980 Oryza sat
C 22	38.2	1.8	190050	1	AL646080	AL646080 Ralstonia
C 23	37.8	1.8	31896	2	AC017879	AC017879 Drosophil
C 24	37.8	1.8	172904	3	AC007414	AC007414 Drosophil
C 25	37.8	1.8	275390	3	AE003831	AE003831 Drosophil
C 26	37.4	1.7	38408	2	AC019786	AC019786 Drosophil
C 27	37.4	1.7	146153	3	AC007452	AC007452 Drosophil
C 28	37.4	1.7	183213	10	AL669952	AL669952 Mouse DNA
C 29	37.4	1.7	195533	2	AC101759	AC101759 Mus muscu
C 30	37.2	1.7	133691	9	AC074347	AC074347 Homo sapi
C 31	37.2	1.7	170398	2	AC013540	AC013540 Homo sapi
C 32	37.2	1.7	329709	1	AP002997	AP002997 Mesorhizo
C 33	37	1.7	82746	1	AF453501	AF453501 Actinosyn
C 34	36.8	1.7	810	6	AX110931	AX110931 Sequence
C 35	36.8	1.7	2016	8	PAN296102	AJ296102 Podospora
C 36	36.8	1.7	10925	1	AE004750	AE004750 Pseudomon
C 37	36.8	1.7	137037	2	AC104321	AC104321 Oryza sat
C 38	36.8	1.7	307150	1	CNSPAX01	AJ248283 Pyrococcu
C 39	36.8	1.7	349980	6	AX041106	AX041106 Sequence
C 40	36.6	1.7	1129	8	AF178952	AF178952 Zea mays
C 41	36.6	1.7	1496	8	CRENOLM	X66412 C.reinhardt
C 42	36.6	1.7	6709	14	MOCDRNAP	L38927 Molluscum c
C 43	36.6	1.7	95727	2	AC091231	AC091231 Homo sapi
C 44	36.6	1.7	170630	2	AC016905	AC016905 Homo sapi
C 45	36.6	1.7	185097	2	AC023148	AC023148 Homo sapi

ALIGNMENTS

RESULT 1  
AE004008/c  
LOCUS  
DEFINITION Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.  
ACCESSION AE004008 AE003849  
VERSION AE004008.1 GI:9106961  
KEYWORDS  
SOURCE Xylella fastidiosa 9a5c.  
Xylella fastidiosa 9a5c  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
Xylella.  
REFERENCE 1 (bases 1 to 10689)  
AUTHORS Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,  
Alvaranga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,  
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,

Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carrer, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A.P., Ferreira, A.J., Ferreira, V.C., Ferito, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., and Marino, C.L.

**TITLE**  
The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis

**JOURNAL**  
NATURE 406 (6792), 151-157 (2000)

**MEDLINE**  
20365717

**PUBMED**  
10910347

**REFERENCE**  
2 (bases 1 to 10689)

**AUTHORS**

Simpson, A.J.G., Reinach, F.C., Artuda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M.C., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R.S., Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carrer, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorry, H., Facincani, A.P., Ferreira, A.J.S., Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F., Lambais, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N., Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L., Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C., Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., de M. Silva, A.J., de Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tshako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M., Meidanis, J., and Setubal, J.C.

**TITLE**  
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

**FEATURES**

**Location/Qualifiers**

1..10689

/organism="Xylella fastidiosa 9a5c"

/db\_xref="taxon:160492"

/clone="9a5c"

/complement(119..367)

/gene="XF1871"

/complement(119..367)

/gene="XF1871"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAF84677.1"

/db\_xref="GI:9106962"

/translation="MTMTLKEDKTEIIDEQRIQIAKLSETAKIAEHEVTHQKL

complement(430..618)

/gene="XF1872"

/complement(430..618)

/gene="XF1872"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAF84678.1"

/db\_xref="GI:9106963"

/translation="MMEEARATAPQGLAASDPORITDAPKKTKGLRKRKYPPTGSASI  
IVLRHVLTORQKKRASP"

941..1288

/gene="XF1873"

941..1288

/note="XF1873"

/note="similar to GI|479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAF84679.1"

/db\_xref="GI:9106964"

/translation="MSKNKIPNQILTNWAGQFQINGKLVTPERTTEWQLRWLSLT  
CTLAREWKMMEARATAPQGLAASDPQRTDALQKTVSHKRHAANETASIIQLRQ  
VLIQROKKRASP"

1351..1662

/gene="XF1874"

1351..1662

/note="XF1874"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAF84680.1"

/db\_xref="GI:9106965"

/translation="MTMTLKEDKTEIIDEQRIQIAKLSETAKIAEHEVTHQTEM  
NAHNEMNARTQENNARTOKLIEETQKVSKESQVPPVVASGLLAAGATAATLFIKL  
H"

complement(1839..2918)

/gene="XF1875"

/complement(1839..2918)

/gene="XF1875"

/note="similar to SPIP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="phage-related protein"

/protein\_id="AAF84681.1"

/db\_xref="GI:9106966"

/translation="MLNLVTGAPGNGKTLAYAVDWLIRQIEIDSLVKSAGVPSRFTYD  
TEGFDVAVRLTGTVVQSAPEDWRTTPQGSVIYIDEAHRMFPAGRCRSDDRVCDL  
DTHRRGGYDLMTQWPTKTHHELRLLVGEHVLNRMAGLQTAGLWRSAQDDPYDI  
HQREKAEVWKFEDRYALYASSTLHTVSHKFRPKKWSALSVCVTCVIGLVFWH  
YSPAILSEASVAGASQASLRAPASLSRSRLVSGMRTYAVLETESAPTLGCV  
SSDRSCRCFNTDGYIDMSVVECRLLASPLFPNVHYAYVTSSSSAAAVSSPSSPS  
SSSVLSLSSVSSAPILGSSVSPSSH"

complement(2918..3202)

/gene="XF1876"

/complement(2918..3202)

/gene="XF1876"

/note="similar to SPIP03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon\_start=1

/transl\_table=11

/product="phage-related protein"

/protein\_id="AAF84682.1"

/db\_xref="GI:9106967"

/translation="MFSWLMLLNILGQTVARGLVAGLALVTVPILVTSALNL  
IVKSGISGADVNLNALLMGFGALSIGSAMLRLALQSLHVGIVKATT"

complement(3195..4412)

/gene="XF1877"

complement(3195..4412)

/note="XF1877"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

```
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84683.1"
/db_xref="GI:9106988"
/translation="MCSRFFLYFISALGLTLVISPFCFLFAQNVPTVEVPTSVRY
SSVDSGVSTAFEARSTALVNGVRYIYTPVDISATLGSIAKAVRRGMAFYNYS
MLSLINGAGWVIDELTHEVMSGPALKEIPVGTVAWMQRPDGHVYFVYTPQGLIAP
INAYQSLSPQPLVSSGSPVSSSRWLYHLEGGGEYIGHLSKTDQSPVDYSSGLPP
SVTLDDGLQVRSPTNVAVLTDSTGAVLLTPELVSAANKLRLEDELKASHAP
DOOPSSGASPPSSSCTANPFCNSWASVCDIDWKFDEFLKKPLVDPDVPYDKL
PQKATWSGJGEGCESPTPIPIEFSGIKTSVEISIQPFCDFAALMRPVVIVATILA
AYIAGGRGVKNV"
gene
complement(4481..4702)
/genes="Xf1878"
CDS
complement(4481..4702)
/genes="Xf1878"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84684.1"
/db_xref="GI:9106969"
/translation="MKSVIDRGTAFFSLISTSLVFAADAAGASFDAGAATAALGAIS
GAVLIGAAKLAPAAISVGWKLGAIFG"
gene
complement(4724..4924)
/genes="Xf1879"
CDS
complement(4724..4924)
/genes="Xf1879"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84685.1"
/db_xref="GI:9106970"
/translation="MAQQLMSLYCTQYDVEARTCSQQAWMVPSSLPIPSYEDVRILL
PHIVMCFVAMGFHFLFTVVRD"
gene
complement(4931..5236)

Query Match          6.0%; Score 129.4; DB 1; Length 10689;
Best Local Similarity 44.9%; Pred. No. 3.1e-23;
Matches 888; Conservative 0; Mismatches 1051; Indels 38; Gaps 9;

Qy 188 CCAATGCGGTGTTATCTTCGGTGGCGGATGACGGGTATCACAGTGTCCGATCAGGGCC 247
Db 10327 CCATGGGGGCGAGTGTCTCTCCGGCAATTGCACCGGAGTCTGGTTTCACCAAGAGAC 10268

Qy 248 TGATCTTTTACCAACACCACTGCGGATACGGTGTCTATCCAGAGCAAGACACGGTGGATC 307
Db 10267 TGGTGATCACTAACCATTCATTCGCTTATGTTGGGATTCAGTTGCACTCTACGGCTTAAA 10208

Qy 308 ACGACTATCTCGCGGATGTTTCTTCGTCAGCATGGGTGAGGAGCTT---CCGATT 363
Db 10207 AGAACTTAATCAAGAGAGGTTCAATGCTTGACTCAGGCTGATGAAGTCAAGTGGGGTC 10148

Qy 364 CCGGGTCTTTCCGTAAGTATCTCCGCAAGATCTGAGGTAAAGGACAGGT--AGAG 421
Db 10147 CGAATGGCGGTATTTAGTGGTGTGAGGAGATTACCGATGTACCGCTCAAGCTAAAGCTG 10088

Qy 422 GACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAGCTCAGGAGGTATGCC 481
Db 10087 CTCGTGGCTGCGGGTANTGACCCCTTCAGCGGTACGACGGCATTTGGAGACATTCAGCA 10028

Qy 482 AAGACTGGGCAAAAAAGAAATCGACGAGAACCAACTCTGCATCGTAGAGCTTTCT 541
Db 10027 AACAGAAATCGCCAAATCGCAGGAGAGCAGGGTTATTCGCTGCCAATTTCTTCAGTTTCG 9968

Qy 542 ATTCACAACAGGATCTCTCTCATGCTCTACGATGTATTCAAGGAGCTTCGTATGGTAT 601
Db 9967 CTGGTGGAAATACCTACAGAGTGTTCAGAACCTTGGAGATCAAGGAGCTGCGTTTGGTTT 9908
```

```
Qy 602 TTGCTCTCCAGCTCTCTAGTAAAGTTTCGAGGCGGATACGGACAACCTGATGTGGCCG 661
Db 9907 ACGCACCGCAGGGGAGTGTGGTAAAGTTTCGGGCGCGAGCTCGATACTGATGTGGCCG 9848

Qy 662 GTACACGGGGAGCTTCAGCGTATTCGCGGTGTATGCGGTGCGGACACCGCGCGCGCG 721
Db 9847 GTCATCCGGTGCCTTTCTTTCTACCGTGCCTACGCTGCTGTAAGAGAGCGGTAAACAG 9788

Qy 722 AATACACGAAGGACAATAAACCCCTATAAGCCCGTTTACTCTCGCTGCGGTATCCATGA 781
Db 9787 GCTTCTCTAAGAAACATACCTTACCGTCCCAACACTGCGTGAATTTAGCGATCAAC 9728

Qy 782 GCTACAAGGCTGACGATGATGCCATGACCATCGGTTCCTCCGGGCGAGTACGGATCGT 841
Db 9727 CATTAGGTGATGGTGACTTCGTGATGCTGCTGCTTACCCAGGAGGTACCAATCTGTTAT 9668

Qy 842 TCACCTCTTGGGGTGTGAAGATCGTATCGAAACGAGACAACTCTCGTATCGAAGTTC 901
Db 9667 CTTTGGTGGCGGAATTTGAGAAATATCTGCTCATCGGACTTATCCGCTGATTTGGCCAG 9608

Qy 902 GCGGTATCAAGCAAGGATCTGGAAGGAGCCATGAGCGGAGATCAGGCTACCGGTATCA 961
Db 9607 TCAAGAAATCTCATTGCTTTGATGAGCGGCTAGCAACAGAAATCCAGACATTCAGTGA 9548

Qy 962 AATATGCGCAAGTATGCTCAGAGTGTCTACTATTTGGAAGAATTCGATCGGTATGAAC 1021
Db 9547 AATATGCGAGCACTTTAGCTGGATTGAATACACATCGAAAACTTTGATCGTCAATGG 9488

Qy 1022 GCGGTCTGCGTCTGTTGACGTGATAGTCTTAAGCGTGGCGGAGAAAGACATTCGCGAG 1081
Db 9487 ATGGTCTCCGAGCGCTTAATGCTATAGTCAAGAACAGAGTGAAGAAACAGACGATTTAG 9428

Qy 1082 ACTGATCCGTAAAGCAAGGAGTGTCTATATGCGGATGTATGTCTTCTCTCGAA 1141
Db 9427 CTTGGTTGAACACAGCGGCGATACGTGGACACGAGCAAGCATTTGGCTGCGCATCAAA 9368

Qy 1142 AGGCTTTAAGGAAGGAGCCAAAGCCGTCAGATGACATTTATTTAGCGGAGAGCGTCT 1201
Db 9367 TTGATCTGACAGAACATACAAAGCAATCAGGATCGTGATTTTGTATTTGGTCAAGTCA 9308

Qy 1202 TCG---GTGGTACCGAGTGTGTTTCTGTTTTCACAGTGTGCGAAGCATTTGCTCAAA 1258
Db 9307 ACGGAATGGTGTCTCATTTGTTAGCGGTGAACCTATATATCGCTTGGCAATCGAGGCA 9248

Qy 1259 CTGATGCTCATGCCGGTATCCTCAAAATCGCTTGAAGCAAGTACAAAGACTACCTCC 1318
Db 9247 AATCTGATGCCCAACGTGAACGGGATATCAGGAGCGGATCTGCCAACCATTTGAAGCA 9188

Qy 1319 CGCTCGACCGTAAAGTGTGCTGCCCGCATGCTCGATATTTGTACGCGGCGTATCCCTGCG 1378
Db 9187 ACTTAAAGCAGATGAGCGACGTTACTTTACCGGAATGATGATCGCGAG---ATGCAGCA 9131

Qy 1379 ACAAGTCCCGGATATATTCAAGAAATGTAATCGACAA-----GAAATTCAAAGGC 1428
Db 9130 ACTGGCTCACCAGATACAAAGCTTCTGTCAAAACAGCGTGTGGCTCAATTTGATGTTT 9071

Qy 1429 GACACGAAGATATGACAGCTTCGTATTTCCACAAGAGTGTGTTTCTTATAGCACAAG 1488
Db 9070 GTTTGGAGATGGCATCCAGCAACTCTCAAGCGCTTGGGTGACACCAAAATTAAG-CAGT 9012

Qy 1489 TTCCATGCTCATGCTCAAGTCCATGGACAAGAAAAGTTTGCCAAAGGCTATTCGAGAAAG 1548
Db 9011 TCTGAGGAAAGCTGAAGTGGTGTAAATGCCCGACCGTGCAGCGCTCGAAGAGCAGTCA 8952

Qy 1549 CCGCAGTAGAGCTTTTCAAGAGCGTAAATAGTCTGCTCGCGTATTTACGCGCCGATCG 1608
Db 8951 CCAGCGATCCGTTATGCTGTGCCCATCATGCCAGCATTTACTGGAGATTTGAACGTCAGA 8892

Qy 1609 ATGCCCAATGCTATGCCATTGAGAGGGGCAAGCGTCTTTTCTTTTCCGCGTTTGGCTG 1668
Db 8891 AAAATCCGTACTGGAGAACTACTTTAAAGCGCTCCGATCTACTTTGCAAGCGCTTAGCTG 8832
```



```
gene
8242..9540
/locus_start=1
/transl_table=11
/product="non-hemolytic phospholipase C"
/protein_id="AA35907.1"
/db_xref="GI:21107164"
/translation="MTSSRRDLKRVAAITAGALPSSIGRALALPANARTGLTRD
LEHVIWMQENRSDHYFGLRGVGFDPRLALQRDGHAVWSQTPDGRVLPFAFD
SONTCAPLKSLSHWSKAGHODPARWAEYDAWVPYKGLTNGYFORHDIYYHALAD
AFTCDGYICSLHGTNPENRMYLFTGTSGPSVNGVAQAVTNADDGNWTADWARKPG
YAALQWYYAQLQAAGWDRVYOBYDNFGNSLAYFHYRDLRDLDDERYLRACVP
GSTAGNAATTQAEHLVAIAADVRANRLPOVSWIPPTAYCEHEPEAPPAYGESLVARL
IDALTANPEVWAKTALLINDENDGCFDHFVPAPLPALDARMGRSNVDTHGEVDPVI
GLGRVPMVLSPTWTRGWSVQFDDHVSFLLERREGVAPENISPMWRRAVSGDLTS
VDFKRPDQSSALSALPSVDDYRARTAAVRDRLPSAPAAATVMPQEQORARALPYA
LQVHARVGEAGAAQVLOFVNSGAFAAFVYNSAAGGPPWYTVLPQLDDAPICATH
EGVALRVHGPNGFLREFAGOVPRASAPSAAPWVEARQEGDALVLEIGNAGORACTVO
LRALDYADPSARTLSLAAGQRETIRLALAASHDWYDLVVEOPGSAFRRLAGHLETR
PSRDPAFGRV"
6617..6766
/locus_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA35908.1"
/db_xref="GI:21107165"
/translation="MRPNLFAVSVLATLVAASQAQAGGMMVPOQLPEIAGPPQOAGL
RLSPR"
6791..7054
/locus_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA35909.1"
/db_xref="GI:21107166"
/translation="MGAVVAGLGGCTASVSPQGLVTVNHHCAYGAIQLNSTAQNLIK
DGFNAVRPADELSSAGPSRIYVLDITDVTAPAKAATPVRR"
7051..7236
/locus_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA35910.1"
/db_xref="GI:21107167"
/translation="MMDAQGLVGLAFDGNWESVSSNWFDPATRMIAVDGRLRWI
MTEVAPQQLKELGVR"
7329..7973
/locus_start=1
/transl_table=11
/product="phosphoglycerate mutase"
/protein_id="AA35911.1"
/db_xref="GI:21107168"
/translation="WRILLRHGETPWAAGRYQGIIDPLSPVCGEQAALGERLHS
LOIDRAVAPSLRAQATAKALGASROALLOTDLQELTAHGEWGLLASEINDKPA
RLRAWREPDVTVLMPGGESLRQVLRNGLMRAADGLGADHTLLVVAHDVAVNRVILC
KILGLPLSLRLSFRQAPTTNLLEGGDDVEHLEVVRLNDCAHHTPFFGEAKHRL"

gene
8242..9540
/locus_start=1
/transl_table=11
/product="folylpolyglutamate synthase"
/protein_id="AA35912.1"
/db_xref="GI:21107169"
/translation="MNATDLSAWLAYIEQQHPSAIAMGLERVRVAVRLQIAAPAKH
VIVVGTTNGKSTVAFIEAIGRAAGKVGAVTSPHLLRYNERNVIRIDGGEASDAQLVDA
FAAVEARGDALTIVFEFGTLAALWLFQSSALELAVLEIGLGRDLDAVNIVSDVAVI

Query Match
Best Local Similarity 56.4%; Score 61.6; DB 1; Length 10811;
Matches 115; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 1936 GGTAACTCCGGTAGCCCGCTATTTCGATAAGACGCCGCTGATCGTCTTTCGAT 1995
DB 7033 GGCACTCCGGTTCGCCGCTGATGGACGCCGCGCAAGCTGCTCGCTGGCTTCGAC 7092
QY 1996 GGCACTGGGAAGCTATGAGTGGTGACATCGATTCGAAACCCGATTCGACGCCACAATC 2055
DB 7093 GGAATTTGGAGTTCGGTGAGCAGCAACTGGATCTTCGACCCGCAATGACCCGCATGC 7152
QY 2056 AGCTGGGACATCGGTACGTTCTTCATGATTCGAAATGGGTGAGTCCGCCCTCTC 2115
DB 7153 GCCGTGGACGGCGCTATCTCGCTTGGATCATGACCGAGGTCTCCGGCGCACAGCTG 7212
QY 2116 ATCCAAGAGCTGAAGTTGATCTAA 2139
DB 7213 TTGAGGAACCTGGCGTCGCTGA 7236

RESULT 3
AF429315/c
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
JOURNAL MEDLINE
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
TITLE Submitted (05-Oct-2001) Psychlatry, Johns Hopkins Medical
JOURNAL Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES
Location/Qualifiers
1..125020
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/notes="Isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
repeat_region
/rpt_type=tandem
```

```
gene
8242..9540
/locus_start=1
/transl_table=11
/product="folylpolyglutamate synthase"
/protein_id="AA35912.1"
/db_xref="GI:21107169"
/translation="MNATDLSAWLAYIEQQHPSAIAMGLERVRVAVRLQIAAPAKH
VIVVGTTNGKSTVAFIEAIGRAAGKVGAVTSPHLLRYNERNVIRIDGGEASDAQLVDA
FAAVEARGDALTIVFEFGTLAALWLFQSSALELAVLEIGLGRDLDAVNIVSDVAVI

Query Match
Best Local Similarity 56.4%; Score 61.6; DB 1; Length 10811;
Matches 115; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 1936 GGTAACTCCGGTAGCCCGCTATTTCGATAAGACGCCGCTGATCGTCTTTCGAT 1995
DB 7033 GGCACTCCGGTTCGCCGCTGATGGACGCCGCGCAAGCTGCTCGCTGGCTTCGAC 7092
QY 1996 GGCACTGGGAAGCTATGAGTGGTGACATCGATTCGAAACCCGATTCGACGCCACAATC 2055
DB 7093 GGAATTTGGAGTTCGGTGAGCAGCAACTGGATCTTCGACCCGCAATGACCCGCATGC 7152
QY 2056 AGCTGGGACATCGGTACGTTCTTCATGATTCGAAATGGGTGAGTCCGCCCTCTC 2115
DB 7153 GCCGTGGACGGCGCTATCTCGCTTGGATCATGACCGAGGTCTCCGGCGCACAGCTG 7212
QY 2116 ATCCAAGAGCTGAAGTTGATCTAA 2139
DB 7213 TTGAGGAACCTGGCGTCGCTGA 7236

RESULT 3
AF429315/c
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
JOURNAL MEDLINE
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
TITLE Submitted (05-Oct-2001) Psychlatry, Johns Hopkins Medical
JOURNAL Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES
Location/Qualifiers
1..125020
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/notes="Isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
repeat_region
/rpt_type=tandem
```

```
/rpt_unit-ctg
complement(<36507. .>36887)
/gene="JPH3"
/notes="JPH3"
mRNA
complement(<36507. .>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507. .36887)
/gene="JPH3"
/notes="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSSGGREFNDDGSGYCGGHEDGKAHGHVCTGPKGOGGYTGWS
HGFEVIGVYWPVSGNTYGTWAGKRGHIGLESKGMVYKGEWTHGFKRGYVRECA
NGAYEGTWSGLDGYGTETYSDG"
BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others
ORIGIN
Query Match 2.08; Score 43.2; DB 9; Length 125020;
Best Local Similarity 9.8%; Pred. No. 5.9;
Matches 46; Conservative 226; Mismatches 194; Indels 4; Gaps 1;
QY 250 ATCTTTACCAACCACTCGCGGATACGGTCTATCCAGAGCCAAAGCACCGTGGATCAC 309
Db 51261 AKCWWSYRYKYCNMMSMAWMAWMMCTAANYIKKTRMRWSKYSSYAYAYAT 51202
QY 310 GACTATCTGCCGATGTTTCGTTTCGCACAGATGGTGAGGAGCTTCGATTCGCGGT 369
Db 51201 MCWYCSMYAARRMGTVSGGKKRWASTSCMCCRCCKHKKGRWCSTYTTGMSGK 51142
QY 370 CTTTCCGTGAAGTATCTCCGCAACATCGTGAAGCTAACGGACAGGTAGAGACAGCTC 429
Db 51141 SRWTSWMACMYWYKGRRKTKSNATSCMRXGAMRSKGMKRYKWSRYKGMWRGWSY 51082
QY 430 AAGGTATCACTGACGAGATGGAGCGCTGCGCAAGCTCAGGAGGTATGCCAAGAACTG 489
Db 51081 KCGSGYSMGMYSGRTSKYCSYTGVCYKCSAKMCKSKSTKSCCTKSKYS---CSMG 51026
QY 490 GCCAAAAGAAATGACAGAGCAACCAACTCTGCATCTGATAGAGCTTCTATTCCAAC 549
Db 51025 RYSSYCTCGWGRMCKMKSSCCASRSCKSYKMGASRCRRGRGMSYWSMSKYSMAS 50966
QY 550 AACGAATCTTCCTCATCTACGATCTATCAAGGAGGTTCGTATGTTGCTCTCT 609
Db 50965 YKSSARGMKCYKTSCKMYIYSHBKSDBSGYKSYRKYBYWTCKMKDSMCSVDSH 50906
QY 610 CCCAGCTCTGTAGTAAGTTCGGAGCGGATACGACAACTGGATGGCCCGCTCACACG 669
Db 50905 MARDWYDKMRKMYSSGTRGYHBHSAWMCMSRSDSGYSGMYHMKSYKWSKSKHMY 50846
QY 670 GGCACCTACCGCTATTCGCGCTGTATCGCGTCCGACACACCGCGCGC 719
Db 50845 SGMKSSVYBMDCMYSYBSKSVBRBRWSGSYBRYKCAVMRRYBHRSS 50796
RESULT 4
AF403408/c 820 bp ds-RNA linear VRL 07-AUG-2002
LOCUS Golden shiner reovirus segment 11 non-structural protein NS3 gene,
DEFINITION complete cds.
ACCESSION AF403408
VERSION AF403408.1 GI:22128470
KEYWORDS Golden shiner reovirus.
SOURCE Golden shiner reovirus.
ORGANISM Viruses; dsRNA viruses; Reoviridae; Aquareovirus.
REFERENCE 1 (bases 1 to 820)
AUTHORS Attoui,H., Fang,Q., Jaafar,F.M., Cantaloube,J.F., Biagini,P., De
Micco,P. and De Lamballerie,X.
TITLE Common evolutionary origin of aquareoviruses and orthoreoviruses
```

```
revealed by genome characterization of Golden shiner reovirus,
Grass carp reovirus, Striped bass reovirus and golden ide reovirus
(genus Aquareovirus, family Reoviridae)
J. Gen. Virol. 83 (pt 8), 1941-1951 (2002)
12124458
JOURNAL PUBMED
REFERENCE 2 (bases 1 to 820)
AUTHORS Attoui,H., de Micco,P. and de Lamballerie,X.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-2001) Virologie, Faculte de Medecine de
Marseille, 27 Boulevard Jean Moulin, Marseille 13005, France
FEATURES
Location/Qualifiers
source
1..820
/organism="Golden shiner reovirus"
/db_xref="taxon:185783"
/notes="Golden shiner reovirus
segment: 11"
43..777
/codon_start=1
/product="non-structural protein NS3"
/protein_id="AAM92755.1"
/db_xref="GI:22128471"
/translation="MATAGITAAIDDLVTSRIDEAITNNTLVORLKKVSSSLKWDYDA
HVSPTAHHELLTPAALADLPPTDDINSFYATALPSVPTSDTLWGTASVEI
IGRAVRIASSDKAPRLILHVTDETIRLEVTFMSLTASPLSTGSLRVTLPRAIIHS
LTKGFDERCWVRLSLTLTKWSTPSTWVTSMTLDTTTLTKPPVTGFNARATVKDSAI
VMFEPADIAITSARRAIVSPMSCVFE"
BASE COUNT 161 a 288 c 196 g 175 t
ORIGIN
Query Match 2.04; Score 42.6; DB 14; Length 820;
Best Local Similarity 54.0%; Pred. No. 4.6;
Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 1958 CGCACAAAACATATGTCGCTATGCCGAGAACCGTCAGTCATATCGCTTCTCTATCG 1917
Db 744 CGCCGACGAGCGGAGGTGCGATGTCGGGAGCGGGGAACACATAACGATGGCGGAGTC 685
QY 1918 AACACGACATACCGGCGGTAACTCCGGTAGCCCGCTATTTCGATAAGAACGCCGCTG 1977
Db 684 TTTAACGTCGACGCGGCTTGATCCAGTCACGCGGGGTTGTCGACACCGCGGTGGT 625
QY 1978 ATCGGTCTTGTTCGATGGCAACTGGGAAGCTATGATGG 2018
Db 624 GTCCAAGCTCCAGTCAGTGTGACCATGCTGCTGGGGTGG 584
RESULT 5
AY051790 1461 bp mRNA linear INV 27-AUG-2001
LOCUS Drosophila melanogaster LD30634 full length cDNA.
DEFINITION AY051790
ACCESSION AY051790
VERSION AY051790.1 GI:15291890
KEYWORDS FLL-CDNA.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1461)
Stapleton,M., Broksstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Farfan,D., Friese,E., George,R.,
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
Direct Submission
Submitted (10-AUG-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
```





Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
**COMMENT** On Feb 17, 2001 this sequence version replaced gi:6984365.

**REFERENCE** Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720  
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

**FEATURES** Location/Qualifiers  
source 1..148847  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="3R"  
/map="85D-85D"  
/clone="BACR32M04 (D969)"  
/clone\_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"  
BASE COUNT 43851 a 31005 c 30227 g 43764 t  
ORIGIN

Query Match 2.0%; Score 42.2; DB 3; Length 148847;  
Best Local Similarity 49.3%; Pred. No. 11;  
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
QY 1470 GGTTCCTTATAGCGACAAAGTTCCATGCTCAAGTCCATGACAGGAAGATTTCG 1529  
DB 7940 GCCCGCTATCAGCCCGCTGCTGCTGCCAGAGTCCAGCGACTCGGTTCTGTGT 7999  
QY 1530 CAAGGCTATCGAGAAGATCGCGAGTAGAGCTTTCCAAAGAGCGTAATAGCTGCTGTCG 1589  
DB 8000 CCAGACCAGGAGAGGACGTGGCAGCATCGCTGCCAAGCGTGATACCATTCGCGCTC 8059  
QY 1590 CGCTATTACGCGCGATGGATGGCAATGCTTATGCCATTGAGAAGGCAAGCGTCTTTT 1649  
DB 8060 CACCACCAATCCCTACAACTGGCGCTCTACTAGCGGACGACTACCGGTGATCTTCAA 8119  
QY 1650 CTTTCCGCTTGGTGAGATGTACCCGGAGCTGCTGCGG 1692  
DB 8120 CATCATCTGTGGTTCATGGTGGTCTTGGGACTGTCTCTGCTG 8162

RESULT 8  
AC008356  
LOCUS AC008356 162593 bp DNA linear INV 06-SEP-2001  
DEFINITION Drosophila melanogaster, chromosome 3R, region 85D-85D, BAC clone BACR43K14, complete sequence.  
ACCESSION AC008356  
VERSION AC008356.5 GI:15451512  
KEYWORDS HTG.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 162593)  
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H., Holt,R.A., Evans,C.A., Gocayne,J.D., Ananidis,P.G., Brandon,R.C., Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busan,D.A.,

Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D., Ferreria,S., Frise,E., Galle,R.E., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J., Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattel,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Swirskas,R., Tector,C., Williams,S.M., Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.  
Sequencing of Drosophila chromosome 3R, region 85D-85D  
Unpublished  
2 (bases 1 to 162593)  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Swirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA  
**COMMENT** On Sep 6, 2001 this sequence version replaced gi:12957666.  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

**FEATURES** Location/Qualifiers  
source 1..162593  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="3R"  
/map="85D-85D"  
/clone="BACR43K14 (D968)"  
/clone\_lib="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"  
BASE COUNT 49906 a 32795 c 32702 g 47190 t  
ORIGIN

Query Match 2.0%; Score 42.2; DB 3; Length 162593;  
Best Local Similarity 49.3%; Pred. No. 11;  
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
QY 1470 GGTTCCTTATAGCGACAAAGTTCCATGCTCAAGTCCATGACAGGAAGATTTCG 1529  
DB 123301 GCCCGCTATCAGCCCGCTGCTGCTGCCAGAGTCCAGCGACTCGGTTCTGTGT 123360  
QY 1530 CAAGGCTATCGAGAAGATCGCGAGTAGAGCTTTCCAAAGAGCGTAATAGCTGCTGTCG 1589  
DB 123361 CCAGACCAGGAGGACGTGGCAGCATCGCTGCCAAGCGTGATACCATTCGCGCTC 123420  
QY 1590 CGCTATTACGCGCGATGGATGGCAATGCTTATGCCATTGAGAAGGCAAGCGTCTTTT 1649  
DB 123421 CACCACCAATCCCTACAACTGGCGGCTCTACTACGGCAGGAGCTACCGGTGATCTTCAA 123480  
QY 1650 CTTTCCGCTTGGTGAGATGTACCCGGAGCTGCTGCGG 1692  
DB 123481 CATCATCTGTGGTTCATGGTGGTCTTCGGGACTGCTCTGCTG 123523

RESULT 9  
AE003682



LOCUS AE003682 224400 bp DNA linear INV 05-OCT-2000  
 DEFINITION Drosophila melanogaster genomic scaffold 14200001386035 section 7  
 of 105, complete sequence.  
 ACCESSION AE003682 AE002708  
 VERSION AE003682.2 GI:10726402  
 KEYWORDS HTG.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 224400)  
 Adams,M.D., Celiker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
 Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,  
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
 Sutton,G.C., Wortman,J.R., Vandeil,M.D., Zhang,Q., Chen,L.X.,  
 Brandon,R.C., Rogers,J.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,  
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
 Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,  
 Andrews-Pfannkoch,C., Balgwin,D., Ballew,R.M., Basu,A.,  
 Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
 Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,  
 Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,  
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
 Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
 Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S.,  
 Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S.,  
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,  
 Guan,P., Harris,M., Harris,N.D., Harvey,D., Helman,T.J.,  
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,  
 Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,  
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,  
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,  
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,  
 McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Moberly,C.,  
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,  
 Muzny,D.M., Nelson,D.L., Nelson,S.M., Moy,M., Murphy,B., Murphy,L.,  
 Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,  
 Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,  
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,  
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,  
 Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,  
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,  
 Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,  
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
 The genome sequence of *Drosophila melanogaster*  
 Science 287 (5461), 2185-2195 (2000)  
 20196006  
 10731132  
 PUBLISHED 10731132  
 REFERENCE 2 (bases 1 to 224400)  
 Adams,M.D., Celiker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.  
 JOURNAL Direct Submission  
 AUTHORS Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
 TITLE Rockville, MD, USA  
 JOURNAL On Oct 9, 2000 this sequence version replaced gi:7299142.  
 COMMENT Location/Qualifiers  
 FEATURES 1. .224400  
 source  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 complement(<235. .>2417)  
 /gene="Dl"  
 /note="CG9745"  
 /map="85D2-85D3"  
 /db\_xref="FLYBASE:FBan0009745"  
 /db\_xref="FLYBASE:FBgn0000412"  
 complement(join(235. .1324,1643. .2221,2355. .2417))  
 /gene="Dl"  
 mRNA

/product="CT27541"  
 /db\_xref="FLYBASE:FBan0009745"  
 /db\_xref="FLYBASE:FBgn0000412"  
 complement(join(672. .1324,1643. .2057))  
 /gene="Dl"  
 /note="Dl gene product"  
 /codon\_start=1  
 /protein\_id="AAF54341.1"  
 /db\_xref="GI:7299143"  
 /db\_xref="FLYBASE:FBan0009745"  
 /db\_xref="FLYBASE:FBgn0000412"  
 /translation="MEEVAVKKRPSKASVGGKSSAAVAASPGLKRRGRPAKNG  
 SSGGQGRPRPKAKTQNDDEDEEDGDGSGAELANNSPSPPTKGRPKSS  
 KGAGSGDSVKTPTGSAKKRAGPKKHQPSSENEEDDDQDNGNSIERRRVGRP  
 SAGVNLNISRGTGLRGPKRRAVESNGDGPQVKKRPPQNKSGSGSTGVVPTG  
 RPRGRANAAPVEKHEDDDDDQDENSEGEEHSSPEKTVVAPKRRGPSLAACKVSK  
 EETPKPSRPAKNIIDDADDADSDAQGHNSKESNEDRAVDGTPTKGDGLKNSDG  
 ENDANGTVSDNYNDSVAA"  
 complement(<3004. .>7907)  
 /gene="CG9746"  
 /map="85D2-85D4"  
 /db\_xref="FLYBASE:FBan0009746"  
 /db\_xref="FLYBASE:FBgn0037663"  
 /db\_xref="FLYBASE:FBgn0009746"  
 complement(join(3004. .3235,3300. .3914,3972. .4605,  
 4986. .5199,5255. .6265,6330. .7367,7561. .7747,7809. .>7907))  
 /gene="CG9746"  
 /product="CT27547"  
 /db\_xref="FLYBASE:FBan0009746"  
 /db\_xref="FLYBASE:FBgn0037663"  
 /db\_xref="FLYBASE:FBgn0009746"  
 complement(join(3004. .3235,3300. .3914,3972. .4605,  
 4986. .5199,5255. .6265,6330. .7367,7561. .7747,7809. .7906))  
 /gene="CG9746"  
 /note="CG9746 gene product"  
 /codon\_start=1  
 /protein\_id="AAF54342.1"  
 /db\_xref="GI:7299144"  
 /db\_xref="FLYBASE:FBan0009746"  
 /db\_xref="FLYBASE:FBgn0037663"  
 /translation="MGNOLVGIAPSQIYAVEHYRSGQFSEIAFSSNMNGSTRFFKVK  
 AKTEGQIVKVFVYHPTPLPLEDKERLEGIKKTLISLANAVNCLPQVRLENDIKAA  
 IMREYKHSIVLVYRSTPLTLEKWTFOILCALQCHQKQKICHGKLENDILITS  
 MSVTHDSIIRLAPSYAGNTLLPAMDIFPSAGCALLEWTEGTAFPELSQLAYRGER  
 LLOOQFTDREEFVIAKRAIVTILIEIGHIKKPAVTEIVQETACYLCHPNLVRHEIC  
 GMIAATTARNLSAIDVQCKIMPAIGALFKAPLIOVEKHILLDCVHPVPVROIFSVLR  
 FODIHFRALARSVRVRSOTRGALPOYEMGOTLRNLFRRLSSEGITDLIEMQLLA  
 GGIVADTRMOLHSAWLNELHGVITTCADPTGSMLATCTSGGKHCTCWLDRFLPI  
 AEIKHPADSWIRKVACHPTPEPSYLLISASQSNNEYSVNIEYTGQRTVLWASPAALSS  
 ASPSDPATTCGILSVGVDSFPFILTGSDDQIRYWDITNPKNSLSKLVPAANDNLADVA  
 FNYGARLIDGQVIEVEQIISMASTGDSQQLRSRNGVTKLWK"  
 complement(<9619. .>13591)  
 /gene="CG8420"  
 /map="85D2-85D4"  
 /db\_xref="FLYBASE:FBan0008420"  
 /db\_xref="FLYBASE:FBgn0037664"  
 join(9619. .10588,10978. .11126,12073. .12352,12809. .13132,  
 13366. .13591)  
 /gene="CG8420"  
 /product="CT24697"  
 /db\_xref="FLYBASE:FBan0008420"  
 /db\_xref="FLYBASE:FBgn0037664"  
 /gene="CG8420"



JOURNAL of nitrogen fixation genes in free-living and symbiotic cells  
MEDLINE J. Bacteriol. 169 (4), 1423-1432 (1987)  
PUBMED 87165745  
FEATURES 2881918

Source Location/Qualifiers  
1..1700  
/organism="Sinorhizobium meliloti"  
/db\_xref="taxon:382"  
<1..255  
/note="ntrB"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA26345.1"  
/db\_xref="GI:152392"  
/translation="EFCVHDNPGVPDPLLPHLFDPFITTKNGSGGLGALVAKIIG  
HGIVCEDSOHSRTTFRVLMPASKGLADEETPMKTGTNG"  
252..1700  
/note="ntrC"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA26346.1"  
/db\_xref="GI:152393"  
/translation="MTGATILVADDDAAIRTVLNOALSAGYDVRLTSNAATLRWITA  
AGDLVVTVDVYMPDENAFEDLLPRIKKARDLPVLVMSAQNTFWTAIKASEKGAAYDL  
LPLMITGESGTGKELRALHSDYKRRNGFPVAINMAAI PRDILIESELFCHKGAFGT  
AQTSTGRFEQAEGGTLFIDEIGDMPDAQTRLLRLOQGEYTTVGGRTP IRSDVRVI  
VKNDKLGSINOGLPRDYLRNVVPLRLPRDRABIDPLRVHFVQOAKKEGLD  
VFREQDELLELMKHPGNVRELNLVRLTALYPQDIVTREIENELSELPDSPI  
EKAAKRSGLSTSQAQVEEMNQIFASGDALPSGLYDRVLAEMEYPLILAALTATRG  
NQIKAADLLGNRNTRLRKIRELGVSVMYRSLA"

BASE COUNT 341 a 553 c 511 g 295 t  
ORIGIN

Query Match 1.9%; Score 41; DB 1; Length 1700;  
Best Local Similarity 55.2%; Pred. No. 14;  
Matches 80; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1654 GCCGGTTTGCGCTGAGATGTACCCGGAGCGTGCTCTGCCGAGCGATGCCAACCTCACCATG 1713  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 681 GCCGCATGCGAGNAATCTACC CGGTGCTTCGCGGGTGATGCAGACCGACTCACGCTG 740  
  
QY 1714 CGTATGAGCTACGGCTCCATCAAGGGATATGAACCGCAGGACGGTGCTGGTACAACATAT 1773  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 741 ATGATCACCGCGAGTCCGCGACCCGCGAAGAACTCGTTGCCCGTGCACTGCACGACTAT 800  
  
QY 1774 CATACGACAGCGAAGGCGGTATTGG 1798  
||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 801 GGCAAGCGGAGAAAATGGCCCTTCG 825

RESULT 12  
SM591787

LOCUS SM591787 329100 bp DNA linear BCT 05-JUL-2002  
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 6/12.  
ACCESSION AL591787 AL591688  
VERSION AL591787.1 GI:15074266  
KEYWORDS  
SOURCE Sinorhizobium meliloti.  
ORGANISM Sinorhizobium meliloti  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Rhizobiaceae; Sinorhizobium.  
1 (bases 1 to 329100)  
Capela,D., Barloy-Hubier,F., Guzy,J., Bothe,G., Ampe,F., Batut,J.,  
Boisard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,  
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,  
Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,  
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.  
Analysis of the chromosome sequence of the legume symbiont  
Sinorhizobium meliloti strain 1021  
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)

TITLE  
JOURNAL 21396507  
MEDLINE  
PUBMED 11481430

```
CDS
2063..2683
/gene="acpD OR SMC01329"
/function="small molecule metabolism; fatty acid biosynthesis"
/note="Product confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE ACYL CARRIER PROTEIN PHOSPHODIESTERASE"
/protein_id="CAC45915.1"
/db_xref="GI:15074269"
/db_xref="SPTREMBL:Q92Q17"
/translation="MKILIHDSGILGSHSVSRRTSIAIVSOLKADRPDAELTYRDLAS
ERVPLHCAQIMAPADLEGVALLAADVRIGRQMLEEFLLAADTVVVGAPMYNFSIQ
LKWIDRLVAGKTFRYTEAGPEGLAKGKLLIVASTRGHYSVAPASAMDHQETYLRS
VFGFEGTIDIEFIRAEGLNLGPDQKQFAIAEAETIAEGNVKLAS"
2792..3076
/gene="SMC01328"
2792..3076
/gene="SMC01328"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAC45916.1"
/db_xref="GI:15074270"
/db_xref="SPTREMBL:Q92KE9"
/translation="MKCVRLRFLALSIAVALVAGSVSAPGGRCRCRRDCAKYE
LGQACIRIGDISYLARCEMNLVTTWKRLQDGCPTAGLAVTIVPSAIP"
complement(3066..3941)
/gene="SMC01327"
complement(3066..3941)
/gene="SMC01327"
/EC_number="2.1.1.-"
/function="macromolecule metabolism; macromolecule synthesis, modification; rna synthesis, modification , dna transcription"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRNA/RRNA METHYLTRANSFERASE PROTEIN"
/protein_id="CAC45917.1"
/db_xref="GI:15074271"
/db_xref="SPTREMBL:Q92Q16"
/translation="MSKDKTGKSDRTHYANLRAHRDARRERGEIPTPRQDKRAKL
SADWKFPALPDQVLLYLGHVTRAAALDNPERRIIRLSVTQNAARLDGLTALPFP
ETVNPQLDRIPDPAIHOGVMTLPPLRKLKLEALDCPLILVLDQVDPHNVGATM
RSVAFADAGALITPMRHSPTESGVLAKSASGALFPIYQITNLADTLEELHELGEWT
IGLDEGPQPLESTFAGDKTALVLGSEKGLRQKTRQTVRKALRLDMPGAIKSLNVSN
AAATAMVAASBHLKAG"
4158..4242
/gene="trNA-TYR_GTA"
4158..4242
/gene="trNA-TYR_GTA"
/product="trNA-Tyr"
/note="codon recognized: UAC; predicted by tRNAscan-SE"
/evidence=not_experimental
4266..4339
/gene="trNA-GLY_TCC"

CDS
4266..4339
/gene="trNA-GLY_TCC"
/product="trNA-Gly"
/note="codon recognized: GGA; predicted by tRNAscan-SE"
/evidence=not_experimental
4602..5777
/gene="tufB OR SMC01326"
4602..5777
/gene="tufB OR SMC01326"
/function="macromolecule metabolism; macromolecule synthesis, modification; proteins and peptides - translation and modification"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE ELONGATION FACTOR TU PROTEIN"
/protein_id="CAC45918.1"
/db_xref="GI:15074272"
/db_xref="SPTREMBL:Q925Y6"
/translation="MAKSKEERNKPHVNICITIGHVDHGKTSLSAATKYEGEKAYDQ
IDAPEKARGITISTAHVEYETPNRHVHVDPGHADYVKNMITGAOMDGAILLVVS
AADGPMPQREHILLARQGVPAIVVFLNKVDQDDAELELLEVELEVEILSYEPFG
DDIPIYKGSALALEDSDKKIGEDAIRELMAAVDAYITPERIDQFPLMIEDVFSI
SGRGTVVTVRVERGIVKVGEEIEIVGIRPTTKTCTGVEMFKLLDQDQAGDNIGALL
RGVDRNGVERGOILCKPQSVKPHRKAEEAYILTKEGGRHTTFFTNYPQFVFRFTD
VTGIVTLPEGTENVMPGDNVTVDELIVPIAMEEKLRFARREGORTVGAGIVASIV"
6125..6790
/gene="SMC01325"
6125..6790
/gene="SMC01325"

Query Match 1.9% Score 41; DB 1; Length 329100;
Best Local Similarity 55.2%; Pred. No. 26;
Matches 80; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Oy 1654 GCCGGTTTCGTGAGATGATACCCCGGACGTCCTGCGCGGCGATGCCAAGTTCACCATG 1713
Db 117496 GCCGCATGTCAGAGAAATCTACCGCGTCTGTCGGCGGTGATGCAGACGCGCTCAGCGTG 117555

Oy 1714 CGTATGACGTACCGCTCATCAAGGATATGAACGCGAGGACGGTGCCTGGTACAACTAT 1773
Db 117556 ATGATCACCAGCGACTCGGCGACCGGCAAGCAACTCGTTGCCCGTGCACGACTAT 117615

Oy 1774 CATACGACAGCAAGGGCGTATTGG 1798
Db 117616 GGCAAGCGGAGAAATGGCCCTTCG 117640

RESULT 13
AF429315 AF429315 125020 bp DNA linear PRI 18-JAN-2002
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll,Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
JOURNAL 21583737
MEDLINE 11694876
PUBMED
REFERENCE 2 (bases 1 to 125020)
```

AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.  
TITLE Direct Submission  
JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical  
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA  
FEATURES Location/Qualifiers  
source 1..125020  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="16q24.3; between D16S520 and WI-12410"  
/note="isolated from a patient with Huntington's  
Disease-like 2 (HDL2)."  
repeat\_region complement(35581..35746)  
/rpt\_type=tandem  
gene /rpt\_unit=ctg  
complement(<36507..>36887)  
/gene="JPH3"  
/note="Jp3"  
mRNA complement(<36507..>36887)  
/gene="JPH3"  
/product="junctophilin 3"  
complement(<36507..>36887)  
/gene="JPH3"  
/note="component of the junctional complex between plasma  
membrane and endoplasmic reticulum"  
/codon\_start=1  
/product="junctophilin 3"  
/protein\_id="AAL40941.1"  
/db\_xref="GI:17646245"  
/translation="MSSGGRFNDGSGYCGWEDKAHGHGVCYTPKPGCEYTGWS  
HGEVLGVYTPWNGNYQGTWAGKRGKRGLESGKRWYKGEWTFGKRGYVRECA  
NGAYIGTWSNGLDGTYGTYSDG"  
BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others  
ORIGIN  
Query Match 1..98; Score 40.4; DB 9; Length 125020;  
Best Local Similarity 9.7%; Pred.No. 33;  
Matches 79; Conservative 349; Mismatches 379; Indels 6; Gaps 2;  
Qy 27 TCTCGGAGCACCCGTGCTTGGTGTCTCAGGGGTAGCCAAAGCCGACAAAGGCATGTG 86  
Db 17023 KCYNSSYWTMSASYSRYSYKWSKMSRSCRSRSCSKSRGCGSGMGKGGKSVGR 17082  
Qy 87 GCTCCTCAACGAATCAATCAGGAGATCTGGATCGAATCGTGAAGTCTGCGCTTTACGCT 146  
Db 17083 KTKRSRGMKGAWYMYRRSRMRKMYSKMYCWCWGRGRCYCSMTSRSAMCC 17142  
Qy 147 CCCGTGGATTGCTCTACAGTTCGACAGCCGCTCCATTCCTCCATCCCGTGGTATCTT 206  
Db 17143 SYCAKCKSWCYSYGMSMGYIACSYRGSMSKSKYCMRGTSTSCGCCCTTTTCCC 17202  
Qy 207 CGGTGGCGGATGTACCGGTATCAGGTGTCCGATCAGGGCTGTATCTTACCAACCA 266  
Db 17203 CCNANTGGG---GAAGCTTTTNCNKKTSYRKRNGCAMCKYNNYNNSSRSCRAMS 17258  
Qy 267 CTGCGGATACCGTCTCAGAGCCAAAGCAGCGTGGATCAGACATCTGCGCGATGG 326  
Db 17259 CTKYKSSMTMSWASCYCMWSMYCYSMSRSMASRGMMSYMKMKMSYCMKCMCMK 17318  
Qy 327 TTTCTTCTCGCAGATGGGTGAGGAGCTCCGATTCGCGCTTCCCGTGAAGTATCT 386  
Db 17319 MCYCMRSMRSGMSYMYASWKSRRRCYCTRCYCMSSKSCYKSYMMRSMRSMKGMK 17378  
Qy 387 CGCAAGATCGTGAAGTGAAGGACAGGTAGAAGGACGATCAAGGTATCACTGACGA 446  
Db 17379 WSRGWSMWSASRSCYKCYK--SMRSMSSKCYRCAGCMKMGYMYRMRSMKRR 17436  
Qy 447 GATGAGCGTCTCGCAAGCTCAGGAGGTATGCCAAGAACTGCCAAAAAGAAATGC 506  
Db 17437 WVGKSAMYRMRMRWKGRRGAMMCMKCYSRMSRCMMMKSYCASCRSAMMSGTMYKCA 17496  
Qy 507 AGACGAGNACCACTCTGCATCTAGAGCCTTTCTTCCCAACAGCAATACTTCTCTCA 566

Db 17497 MCMSSYNCTCWTCTSYMYRRTCMCKGWSSYTKSKSSSSSSSYKKGRRKSYSMCCTSRG 17556  
Qy 567 CGTCTACGATGATTTCAAGGACGTTGCTATGCTATGCTATGCTCTCCAGCTCTGTAGTAA 626  
Db 17557 AMSWRRCCYMRGASMRAGSMRRRAKRSWGGRSKWWMTGGMWRSKYYYCTCRRMM 17616  
Qy 627 GTTCGGAGCGGATACGACAACTGATGTGCGCGCTCACACGGCGGACTTCAGCGTATT 686  
Db 17617 TYMCCWRRRSYMYRYSMSAMGRKSSWSGMRMGSASSRRCKSASRSSWCSRRMKMRG 17676  
Qy 687 CCGCGTGTATCCGCTCGGACAAACCGCGCGCGGCAATACAGCAAGGACCAATAAACCTA 746  
Db 17677 SCWSKMWGSRSSASCKGKSGMRRRRSRYSRKYRGRGKRRKSMFKSGSGSKSKCW 17736  
Qy 747 TAAGCCGCTTTTACTTCGCTGCCGTATCCATCAAGAGCTTACAGGCTGACGACTATGCCAT 806  
Db 17737 KRSGMTSSCYYSASSCMWMMSSKSCCMCMKRRACCCYSSNSCTSMYRCCWGGNKS 17796  
Qy 807 GACCATCGTTCGCGGCGGACGATCGGATCGTAT 839  
Db 17797 YSTCGTTCCKCTGYKKRTWYWMYSWTSMYM 17829  
RESULT 14  
AF051784  
LOCUS 4314 bp mRNA linear VRT 18-JUL-1998  
DEFINITION xenopus laevis l4s cohesin SMC1 subunit mRNA, complete cds.  
ACCESSION AF051784  
VERSION AF051784.1 GI:3328230  
KEYWORDS .  
SOURCE Xenopus laevis.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
Xenopodinae; Xenopus.  
1 (bases 1 to 4314)  
Losada,A., Hirano,M. and Hirano,T.  
Identification of Xenopus SMC protein complexes required for sister  
chromatid cohesion  
Genes Dev. 12 (13), 1986-1997 (1998)  
98315077  
9649503  
2 (bases 1 to 4314)  
Losada,A., Hirano,M. and Hirano,T.  
Direct Submission  
Submitted (03-MAR-1998) Cold Spring Harbor Laboratory, 1 Bungtown  
Road, Cold Spring Harbor, NY 11724, USA  
JOURNAL Location/Qualifiers  
1..4314  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
158..3856  
/note="member of structural maintenance of chromosomes  
family of ATPases; SMC protein"  
/codon\_start=1  
/product="l4s cohesin SMC1 subunit"  
/protein\_id="AAC26807.1"  
/db\_xref="GI:3328231"  
/translation="MGFLKLIETFNKSYKGRQIIGPFHRTAIIIGPNSGSKSLMDA  
ISVFLGKTSNLRVKTLDLHGPVCPAANPAFVSMYVSEDSCEKVSFVIVGSS  
SEYKINKYVOLSEYSDSLKGLIKARPLVPGAVESIAMNPKPRTALPEISR  
AQLOQLFKLYHNSEIEKLNKLSVKNKEDKDKHVEELKDKKYLKGMRE  
QAEKEIKEDAEINOKLPQYIKAKENSHKIKFPAKKSQNAOKYKRRKAMD  
ELKEMLSVEKARQEPERMEESOSOGRODLTFENOVKKYHRLKEASRAATLAQ  
LEFNRRQKADQRLDLERKKVETEAKIKQKLELEENKRIEKLLEYIATSKQSL  
EQNLEETFEVEMAKRIDEINSELNQVMDARIDQESSROQRKAEINIESIK  
RYPGYSYVGLIDLCQPTKKYQIATVTKVQNMDAIIVDSKTRDRCIOYIKQOR  
PLDYLPLDYLEVPTDERLREKAGLIVIRYEPPIKALQYACGNALVCDNVED  
ARITAFGHORHKTVALDGLTFQKSGYVSGASDLKAKARRDEKAVDKLKEKRLT  
EELKEQMKARKKAEELQVQSOAHGLQMLRLKYSOSDLEQYKTRILANMQEKSLESE  
LANFSPRINDIKRLIQDREMKDLKEMKQVEDEVPEEFCEIGVIRNIRREFEEKVK  
RQNETAKKRLFEFNQKTRGLIQDYKKNQKEDQGVQVTEQSVYKQKDDNETELKKEE

ORHMKTIIDETMAQLDLKNOHLAKKSEVNDKNHLMEDIRKKLSANKEVTHLQKEVTA  
 IETKLEQKSDRNLQACKMSDIKPLSKGTMDDISQEEGSSQGEESASSORSSTV  
 YAKEALTEIDYDLSDELQDAVADDIKOEMSALHOKINEQQSILQRIAPNNKAMEK  
 LESVRDKFORTSEFAARKRAKAKAQAEPOTKKERDFNACFSVATNIDEIYKAL  
 SRNSQAQFLGPNPEPYLDGYNVCYAPGKFRPMDNLSSGKETVAALALFLAHS  
 YKSPFVLDEIDAAIDNTNIGKVANTIKREQSMNFQAIVISLKEEFYKAESLIGVY  
 PEOGDCVISKVLFDLTKYDPDANPND"

BASE COUNT 1408 a 821 c 1141 g 944 t

ORIGIN

Query Match 1.9%; Score 40; DB 5; Length 4314;  
 Best Local Similarity 58.3%; Pred. No. 28;  
 Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 399 GAAGGTAAAGCTCAGGAGTATGCCAAGACTGGCCCAAGAAATATCAGACGAGTGGAGCGTCT 458

Db 916 GAACATATGCAAGGTGGAGGAGAGCTCAAGCAAGAGAAAGATTTGGGAAGAT 975

QY 459 GCGCAAGCTCAGGAGTATGCCAAGACTGGCCCAAGAAATATCAGACGAGAAACCA 518

Db 976 GATGAGAGCAGCAGCGCCATTGAGAAAGAGATCAAGAGAGGATGAGAACTGAATCA 1035

RESULT 15

AC121850 202306 bp DNA linear HTG 21-MAY-2002  
 LOCUS Mus musculus chromosome UNK clone RP24-92E18, WORKING DRAFT  
 DEFINITION SEQUENCE, 11 unordered pieces.

ACCESSION AC121850

VERSION AC121850.1 GI:21039949

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE Mus musculus.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 202306)

McPherson,J.D. and Waterston,R.H.

TITLE The sequence of Mus musculus clone

Unpublished

2 (bases 1 to 202306)

McPherson,J.D. and Waterston,R.H.

Direct Submission

Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park

Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)

----- Project Information -----

Center project name: M.BB0092E18

----- Summary Statistics -----

Sequencing vector: M13; 0%

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 197589 bases at least Q40

Consensus quality: 198358 bases at least Q30

Consensus quality: 198775 bases at least Q20

Insert size: 172000; agarose-fp

Insert size: 201306; sum-of-contigs

Quality coverage: 21.02 in Q20 bases; agarose-fp

Quality coverage: 11.20 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

consists of 11 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
 \* be preserved.

1 1106: contig of 1106 bp in length

\* 1107: gap of unknown length

\* 1207: contig of 1077 bp in length

\* 2284: gap of unknown length

\* 2384: contig of 1668 bp in length

\* 4052: gap of unknown length

\* 4151: contig of 1279 bp in length

\* 5431: gap of unknown length

\* 5531: contig of 1017 bp in length

\* 6548: gap of unknown length

\* 8793: contig of 2145 bp in length

\* 8893: gap of unknown length

\* 21576: contig of 12683 bp in length

\* 21676: gap of unknown length

\* 43695: contig of 22020 bp in length

\* 43795: gap of unknown length

\* 43796: contig of 51659 bp in length

\* 95455: gap of unknown length

\* 95555: contig of 49909 bp in length

\* 145464: gap of unknown length

\* 145564: contig of 56743 bp in length.

FEATURES

source

1. 202306

/organism="Mus musculus"

/db\_xref="taxon:10090"

/chromosome="UNK"

/clone="RP24-92E18"

1. 1106

/note="assembly\_name:Contig25"

1207. 2283

/note="assembly\_name:Contig28"

2384. 4051

/note="assembly\_name:Contig32"

4152. 5430

/note="assembly\_name:Contig34"

5531. 6547

/note="assembly\_name:Contig35"

6648. 8792

/note="assembly\_name:Contig36"

clone\_end:SP6

vector\_side:right

8893. 21575

/note="assembly\_name:Contig37"

21676. 43695

/note="assembly\_name:Contig38"

43796. 95454

/note="assembly\_name:Contig39"

95555. 145463

/note="assembly\_name:Contig40"

145564. 202306

/note="assembly\_name:Contig41"

BASE COUNT 55798 a 44381 c 43982 g 57115 t 1030 others

ORIGIN

Query Match 1.9%; Score 40; DB 2; Length 202306;

Best Local Similarity 52.4%; Pred. No. 45;

Matches 88; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 1444 GCAGACTTCGATTCGACAAAGAGTGGTTCTTATACGACAAAGTCCATGCCATGCTC 1503

Db 147941 GTACCCCTTCAACTCTGACATAGTATGTATCTTATGTTGGGGTGGCCATCTTA 148000

Qy 1504 AAGTCCATGGCAAGAAAAGTTTCCCAAGGCTATCGAGAAGATCCCGCCAGTAGAGCTT 1563

Db 148001 ATCCCATGAGATGAGCTAACGTAAAGATAGGCAACACAGGAGACAGAGCTG 148060

Qy 1564 TCCAGAGCGTAATAGTCTGCTCGCGCTATTTTCAGGCCGATGCGATG 1611

Db 148061 TCAAGGAGCCCTTCTAGCTTCTGTTGGAATGATTTCCATCTGGAGCTCTG 148108

Search completed: May 23, 2003, 05:02:07  
Job time : 5341 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 02:20:11 ; Search time 327 Seconds  
(without alignments)  
14730.961 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaatgaataaaag.....aagactgaattgatctaa 2139

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2139	100.0	2139	24	AA143635
2	42.2	2.0	1147	23	AB114089
3	42.2	2.0	3632	23	AB114088
4	40.6	1.9	2868	23	AB121554
5	38.8	1.8	4013	23	AB118653
6	38.6	1.8	3621	22	AA148797
7	38.6	1.8	4848	22	AA148798
8	37.8	1.8	10757	23	AB121478
9	37.4	1.7	9845	23	AB118652

10	37	1.7	786	22	AA143635
11	37	1.7	1077	22	AA143635
12	36.8	1.7	810	22	AA143635
13	36.8	1.7	349980	22	AA143635
14	36.4	1.7	1090	21	AA143635
15	36.4	1.7	1292	22	AA143635
16	36.4	1.7	1649	21	AA143635
17	36.4	1.7	1768	22	AA143635
18	36.2	1.7	345	22	AA143635
19	36.2	1.7	536	21	AA143635
20	36.2	1.7	2280	17	AA143635
21	36.2	1.7	2281	24	AA143635
22	36.2	1.7	349980	22	AA143635
23	36	1.7	292	24	AA143635
24	35.6	1.7	6174	22	AA143635
25	35.6	1.7	8617	22	AA143635
26	35.6	1.7	32313	22	AA143635
27	35.2	1.6	449	20	AA143635
28	35.2	1.6	4590	22	AA143635
29	35	1.6	1291	24	AA143635
30	35	1.6	1910	23	AA143635
31	34.8	1.6	632	21	AA143635
32	34.8	1.6	1368	24	AA143635
33	34.8	1.6	5059	20	AA143635
34	34.6	1.6	639	21	AA143635
35	34.6	1.6	672	24	AA143635
36	34.6	1.6	975	23	AA143635
37	34.6	1.6	2975	23	AA143635
38	34.6	1.6	5133	23	AA143635
39	34.4	1.6	1163020	24	AA143635
40	34.2	1.6	3059	8	AA143635
41	34	1.6	2759	23	AA143635
42	34	1.6	3972	24	AA143635
43	34	1.6	5837	12	AA143635
44	34	1.6	5837	14	AA143635
45	34	1.6	5837	14	AA143635

ALIGNMENTS

RESULT 1  
AA143635  
ID AA143635 standard; DNA; 2139 BP.

AC AA143635;

DT 05-SEP-2002 (first entry)

Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.

Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;

DPP-7 inhibitor identification; periodontal disease; gingivitis;

periodontitis.

Porphyromonas gingivalis.

Key Location/Qualifiers

CDS 1..2139

FT /\*tag= a

FT /product= "Porphyromonas gingivalis DPP-7"

PN W0200238742-A2.

XX 16-MAY-2002.

XX 08-NOV-2001; 2001WO-US46782.

XX 08-NOV-2000; 2000US-246827P.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

XX .Travis J, Potempa JS, Banbula A, Bugno M;

Merozoite surface  
Merozoite surface  
Fonsecaea petrosoli  
Pyrococcus abyssi  
Human kidney disea  
Human polynucleoti  
Human kidney-speci  
Human polynucleoti  
C glutamicum codin  
Mouse secreted exp  
Mouse neurotrophin  
NPY-R gene. Mus  
C glutamicum codin  
Corn tassell-derive  
Human immune/haema  
Human immune/haema  
Human immune/haema  
DNA encoding a hum  
Yeast AOD9604-asso  
Arabidopsis cDNA e  
DNA encoding novel  
Fusarium venenatum  
Streptococcus poly  
Stealth virus nucl  
Aspergillus oryzae  
M. capsulatus gene  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Listeria innocua c  
Sequence encoding  
Drosophila melanog  
Neisseria meningit  
Polymerase gene.  
Thermococcus litor  
T. litoralis DNA p

XX WPI; 2002-490075/52.  
 DR P-PSDB; AA015205.  
 XX  
 PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
 the dipeptidylpeptidase for protecting an animal from periodontal  
 PT disease caused by Porphyromonas gingivalis -  
 XX  
 PS Claim 11; Fig 4; 65pp; English.  
 XX  
 CC The invention comprises the amino acid and coding sequence of the  
 CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7  
 CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
 CC second and third amino acids from the N-terminal end of a target peptide.  
 CC The DPP-7 target peptide has an aliphatic or aromatic residue as a  
 CC substituent on the alpha-carbon atom of the second amino acid from the  
 CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are  
 CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for  
 CC reducing the growth of a bacterium and protecting an animal from a  
 CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
 CC or periodontitis). The present DNA sequence encodes the Porphyromonas  
 CC gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the invention.  
 XX  
 S0 Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other;

Query Match 100.0%; Score 2139; DB 24; Length 2139;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAAAATGAAATTAATAAGATTTCTTCGAGAGAGCCCTGCTGTGGGTGCTTCAGGG 60  
 Db 1 ATGCAAAATGAAATTAATAAGATTTCTTCGAGAGAGCCCTGCTGTGGGTGCTTCAGGG 60

Qy 61 GTAGCAAAAGCGCAAAAGGATGTGGCTCTCAACGAACCAATCAGAGAAATCTGGAT 120  
 Db 61 GTAGCAAAAGCGCAAAAGGATGTGGCTCTCAACGAACCAATCAGAGAAATCTGGAT 120

Qy 121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTGCGCTCTACAGTTTGCACAAAGCCG 180  
 Db 121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTGCGCTCTACAGTTTGCACAAAGCCG 180

Qy 181 TCCATTGCCAATGCGGTGGTTATCTTTCGGTGGCGGATGTACGGGTATCACAGTGTCCGAT 240  
 Db 181 TCCATTGCCAATGCGGTGGTTATCTTTCGGTGGCGGATGTACGGGTATCACAGTGTCCGAT 240

Qy 241 CAGGCGCTGATCTTTACCAACCAACACCTGCGGATACGCTGTATCCAGAGCAAGACGACG 300  
 Db 241 CAGGCGCTGATCTTTACCAACCAACACCTGCGGATACGCTGTATCCAGAGCAAGACGACG 300

Qy 301 GTGGATCAGACTATCTCGGGGATGGTTTCGTTCTCGACGATGGGTGAGGAGCTTCCG 360  
 Db 301 GTGGATCAGACTATCTCGGGGATGGTTTCGTTCTCGACGATGGGTGAGGAGCTTCCG 360

Qy 361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAAACGACAAAGTAGAA 420  
 Db 361 ATTCCGGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAAACGACAAAGTAGAA 420

Qy 421 GGACAGCTCAAGGGTATCACGTACGAGATGGAGCGTCTGCCCAAGCTCAGAGATATGC 480  
 Db 421 GGACAGCTCAAGGGTATCACGTACGAGATGGAGCGTCTGCCCAAGCTCAGAGATATGC 480

Qy 481 CAAGAATCGGCAAAAGAAATGACAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540  
 Db 481 CAAGAATCGGCAAAAGAAATGACAGACGAGAACCAACTCTGCATCGTAGAGCCTTTC 540

Qy 541 TATTCACACAGGAATCTCTCATCTGCTACGATGTATTCAGGAGGCTTCGTATGGTA 600  
 Db 541 TATTCACACAGGAATCTCTCATCTGCTACGATGTATTCAGGAGGCTTCGTATGGTA 600

Qy 601 TTTGCTCTCCAGCTCTGTAGTAAAGTTCCGAGCGATACGACAACTGATGTGGCCG 660  
 Db 601 TTTGCTCTCCAGCTCTGTAGTAAAGTTCCGAGCGATACGACAACTGATGTGGCCG 660

Qy 661 CGTCACACGGCGGACTTCAGCGTATTCGCGGTGTATGCGGTGCGCAACACGCGCGCC 720  
 Db 661 CGTCACACGGCGGACTTCAGCGTATTCGCGGTGTATGCGGTGCGCAACACGCGCGCC 720

Qy 721 CAATACACAGGACAAATAAACCCCTATAAGCCGTTTACTTTCGCTGCGGTATCCATGCAA 780  
 Db 721 CAATACACAGGACAAATAAACCCCTATAAGCCGTTTACTTTCGCTGCGGTATCCATGCAA 780

Qy 781 GGCTACAAGGCTGACGACTATGCCATGACCATCGCTTTCCCGGGCAGTACGGATCGCTAC 840  
 Db 781 GGCTACAAGGCTGACGACTATGCCATGACCATCGCTTTCCCGGGCAGTACGGATCGCTAC 840

Qy 841 CTCACCTTCTTGGGTGTGGAAGATCGTATCGAAACGAGAAACAATCTCTCGTATCGAAGTT 900  
 Db 841 CTCACCTTCTTGGGTGTGGAAGATCGTATCGAAACGAGAAACAATCTCTCGTATCGAAGTT 900

Qy 901 CGCGGTATCAAGCAAGGATCTGGAAGAACGCAATGAGCGGAGATCAGGCTACCGGTATC 960  
 Db 901 CGCGGTATCAAGCAAGGATCTGGAAGAACGCAATGAGCGGAGATCAGGCTACCGGTATC 960

Qy 961 AAATATGCCACCAAGTATGCTTCAGAGTCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020  
 Db 961 AAATATGCCACCAAGTATGCTTCAGAGTCTAACTATTGGAAGAATTCGATCGGTATGAAC 1020

Qy 1021 CGCGTCTCGCTCTTTCAGCTGATAGGTGCTGAAGCGTCCGAGGAAAGAGCATTCGCA 1080  
 Db 1021 CGCGTCTCGCTCTTTCAGCTGATAGGTGCTGAAGCGTCCGAGGAAAGAGCATTCGCA 1080

Qy 1081 GACTGGATCCGTATAGACGCGCAAGAGTCTGTCTATGCGGATGTATTTCTCTCTCGAA 1140  
 Db 1081 GACTGGATCCGTATAGACGCGCAAGAGTCTGTCTATGCGGATGTATTTCTCTCTCGAA 1140

Qy 1141 AAGCTTATAGGAGGAGCCAAAGCCCAAGCTGAGATGACTTATTTGAGCGAGACGCTC 1200  
 Db 1141 AAGCTTATAGGAGGAGCCAAAGCCCAAGCTGAGATGACTTATTTGAGCGAGACGCTC 1200

Qy 1201 TTCGTGTGTACGAGGTGTTTTCGTCACAGTTTTCGCAACGATTTGGCTACAAATCCT 1260  
 Db 1201 TTCGTGTGTACGAGGTGTTTTCGTCACAGTTTTCGCAACGATTTGGCTACAAATCCT 1260

Qy 1261 GATGCTCATGCGGTATCTCAATCGCTTTCAGCAAGTACAAAGAGTACCTCCCTCG 1320  
 Db 1261 GATGCTCATGCGGTATCTCAATCGCTTTCAGCAAGTACAAAGAGTACCTCCCTCG 1320

Qy 1321 CTCGACCGTAAAGTGTCTGCGCGCATGCTCGATATTGTACCGCGGTATCCCTCGCCAC 1380  
 Db 1321 CTCGACCGTAAAGTGTCTGCGCGCATGCTCGATATTGTACCGCGGTATCCCTCGCCAC 1380

Qy 1381 AAGCTCCCGGATATATTCAGAATGTATTCAGAAGAAATTCAGAAGGACACGAGAAAG 1440  
 Db 1381 AAGCTCCCGGATATATTCAGAATGTATTCAGAAGAAATTCAGAAGGACACGAGAAAG 1440

Qy 1441 TATCGAGACTTCGTATTCGACAAAGTGTGGTTCCTTATAGCGACAAGTTCATGCCATG 1500  
 Db 1441 TATCGAGACTTCGTATTCGACAAAGTGTGGTTCCTTATAGCGACAAGTTCATGCCATG 1500

Qy 1501 CTCGAAGTCCATGGCAAGGAAAGTTTTCGCAAGGCTATCGAGAAGATCCGCGCAGTAG 1560  
 Db 1501 CTCGAAGTCCATGGCAAGGAAAGTTTTCGCAAGGCTATCGAGAAGATCCGCGCAGTAG 1560

Qy 1561 CTTTCCAAAGAGCGTAAATAGCTGCTCGCGCTATTCAGGCGGATTCGATGGCAATGCC 1620  
 Db 1561 CTTTCCAAAGAGCGTAAATAGCTGCTCGCGCTATTCAGGCGGATTCGATGGCAATGCC 1620

Qy 1621 TATGCCATTGAGAGGCAAGCGCTCTTTTTCGCGGTTTTCGCTGAGATGTACCCCGGA 1680  
 Db 1621 TATGCCATTGAGAGGCAAGCGCTCTTTTTCGCGGTTTTCGCTGAGATGTACCCCGGA 1680

Qy 1681 CGTGTCTGCGCGAGCGATGCCAACTTCACTACCTATGAGTACGGCTTCCATCAAGGA 1740  
 Db 1681 CGTGTCTGCGCGAGCGATGCCAACTTCACTACCTATGAGTACGGCTTCCATCAAGGA 1740

Qy 1741 TATCAACCGCAGGAGCGGTGCTGTGTACAACTATCATACGACAGGCAAGGGGTATTGGA 1800





```
Query Match      1.8%; Score 38.8; DB 23; Length 4013;
Best Local Similarity 47.2%; Pred. No. 2.1;
Matches 118; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 1294 GACAAGTACAAAGACTACCTCCCTCGCTCGACCGTAAGTGCTGCCCGCATGCTCGAT 1353
DB 222 GACAAATAACGGATGGCTCATTTGGATGGTCGGTAATATATCCGGCTCGCATGCC 281
QY 1354 ATTGTAGCCGGCGGTATCCCTGCGCACAAGCTCCCGCATATATTCAAGAATGTAATGCAC 1413
DB 282 ATGGGTACAGCGTTGTGGCTACGACACCGGGGACCATCCACGGCAGTAACATCCAT 341
QY 1414 AGAAATTCAAAGCGCACAGAAAGTATGCAGACTTCGTATTCGACAGAGTGTGGTT 1473
DB 342 CGAATCGTCTTCCTCGGCTTCAAGCAGTATCTGGAGCTCGATTTCGACGAGACTTTTCGTT 401
QY 1474 CCTATAGCCACAGATTCATGCTCAAGTCCATGTCACAGTACACAGGAAAAGTTGCCAAG 1533
DB 402 CCGAGGGCGGAGAAAGATGACGAGGACTTCACCTCGTCCGCTGGGTACGATGCAGAT 461
QY 1534 GCTATCGAGA 1543
DB 462 GATGGGGAGA 471

RESULT 6
ID AAH48797/c
XX AAH48797 standard; DNA; 3621 BP.
AC AAH48797;
XX
XX 05-NOV-2001 (first entry)
DT
DE P. pantotrophus GBI7 DNA encoding ORF2, ORF3 and soxXYZA.
XX
XX Sulphur oxidation; sox; ORF1; ORF2; ORF3; soxX; soxY; soxZ; soxA;
KW reduced sulphur; thiosulphate; transgenic plant; transgenic animal; ds.
XX
XX Paracoccus pantotrophus.
OS
XX
FH Key Location/Qualifiers
FT CDS 1..738
FT /tag= a
FT /transl_except= (pos:643..645,aa:Leu)
FT /product= "ORF2"
FT 752..1312
FT /tag= b
FT /product= "ORF3"
FT 1427..1900
FT /tag= c
FT /product= "soxX"
FT 1941..2363
FT /tag= d
FT /product= "soxY"
FT 2385..2714
FT /tag= e
FT /product= "soxZ"
FT 2749..3621
FT /tag= f
FT /product= "soxA"
XX
XX DE10006201-A1.
PN
XX
XX 16-AUG-2001.
PD
XX
XX 11-FEB-2000; 2000DE-1006201.
XX
XX 11-FEB-2000; 2000DE-1006201.
XX
XX (FRIE/) FRIEDRICH C.
XX
XX Friedrich C;
XX
XX WPI; 2001-497622/55.
XX
XX New genes involved in oxidation of reduced sulphur compounds,
PT particularly conversion of thiosulphate to sulphate, and related
```

```
XX
DR WPI; 2001-497622/55.
DR P-PSDB; AAB86529, AAB86530, AAB86531, AAB86532, AAB86533, AAB86534.
XX
XX New genes involved in oxidation of reduced sulphur compounds,
PT particularly conversion of thiosulphate to sulphate, and related
PT proteins, from Paracoccus pantotrophus
XX
XX Claim 1; Page 10-14; 22pp; German.
XX
XX This invention describes novel genes (I) of the sox (sulphur oxidation)
CC region designated ORF1, ORF2, ORF3 (Open Reading Frame), and soxXYZA
CC from Paracoccus pantotrophus GBI7. The proteins (II) encoded by (I) are
CC used for oxidation of reduced sulphur compounds in cellular or cell-free
CC systems, particularly of thiosulphate to sulphate. (I) is also used for
CC expression of the proteins and to prepare transgenic plants or animals
CC or transformed microorganisms. This sequence encodes the ORF2, ORF3 and
CC the soxXYZA proteins from the P. pantotrophus sox region.
XX
XX Sequence 3621 BP; 650 A; 1153 C; 1186 G; 632 T; 0 other;

Query Match      1.8%; Score 38.6; DB 22; Length 3621;
Best Local Similarity 48.8%; Pred. No. 2.3;
Matches 104; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1847 TCGACCTCTTCCGCAACCAAACTATGTCGCTATGCGGAGACGGTCAGCTCCATATCG 1906
DB 3026 TCGAGCGCGCGATCCACGAGACCATGCGGGGTGTGGAATCGTCCCGCTCAAGATCC 2967
QY 1907 CTTTCCTATCGAACACGACATCAGCGGGGTAACTCCGGTAGCCCGGTATTCGATAGA 1966
DB 2966 CGCGTGTCTGTCGCGAAAGTGCCAGCCGGAATAGATCGTGTGGAACGATATCGGCCAGA 2907
QY 1967 ACGGCGCTGATCGGTCTTTCGATGCGCAACTGGAACTATGAGTGGTCACATCG 2026
DB 2906 AAGCGGGCGGTGCGGTCTTGGTCACGATCTCGACCGGTCCGCTGCTGGTTTCGATGACC 2847
QY 2027 AGTTCGAACCCGATCTCGAGCGCAATACGACG 2059
DB 2846 AGCCCGTCTCGACCGGTCCGCGAGGCCGCG 2814

RESULT 7
AAH48798/c
ID AAH48798 standard; DNA; 4848 BP.
XX
XX AAH48798;
XX
XX 05-NOV-2001 (first entry)
DT
DE P. pantotrophus GBI7 DNA encoding sox-associated proteins.
XX
XX Sulphur oxidation; sox; ORF1; ORF2; ORF3; soxX; soxY; soxZ; soxA;
KW reduced sulphur; thiosulphate; transgenic plant; transgenic animal; ds.
XX
XX Paracoccus pantotrophus.
OS
XX
XX DE10006201-A1.
PN
XX
XX 16-AUG-2001.
PD
XX
XX 11-FEB-2000; 2000DE-1006201.
XX
XX 11-FEB-2000; 2000DE-1006201.
XX
XX (FRIE/) FRIEDRICH C.
XX
XX Friedrich C;
XX
XX WPI; 2001-497622/55.
XX
XX New genes involved in oxidation of reduced sulphur compounds,
PT particularly conversion of thiosulphate to sulphate, and related
```

PT proteins, from Paracoccus pantotrophus -  
 PS Claim 1; Fig 1; 22pp; German.  
 XX  
 CC This invention describes novel genes (I) of the sox (sulphur oxidation)  
 CC region designated ORF1, ORF2, ORF3 (Open Reading Frame), and soxXYZA  
 CC from Paracoccus pantotrophus GB17. The proteins (II) encoded by (I) are  
 CC used for oxidation of reduced sulphur compounds in cellular or cell-free  
 CC systems, particularly of thiosulphate to sulphate. (I) is also used for  
 CC expression of the proteins and to prepare transgenic plants or animals  
 CC or transformed microorganisms. This sequence encodes the P. pantotrophus  
 CC sox region associated proteins described in the invention.  
 XX  
 SQ Sequence 4848 BP; 885 A; 1584 C; 1552 G; 827 T; 0 other;  
 Query Match 1.8%; Score 38.6; DB 22; Length 4848;  
 Best Local Similarity 48.8%; Pred. No. 2.7;  
 Matches 104; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
 QY 1847 TCGACCTCTTCCCGCACCAAAACATATGTCGCTATGCCGAGACGGTCAGCTCCATATCG 1906  
 DB 4253 TCGAGCGCGGATCCACGAGACCATGCGGGGTTGTGGAATCGTCCGCTCAAGATCC 4194  
 QY 1907 CTTTCCTATCGAACACGACATCACGGGCGGTAACTCCGGTAGCCCCGCTATTCGATAAGA 1966  
 DB 4193 CGCGTGCTGTGTCGCGAAAGTGCACGCGGAATAGATCGTGTGGAACGATATCGGCCAGA 4134  
 QY 1967 ACGCGCTGTGATCGGTCTTTCATGCGCAACTGGAGGCTATGAGTGGTGACATCG 2026  
 DB 4133 AAGCGGGCGGTGCGGTCTTGTGCACGATCTGACCGGTCCGCTGCGTTCGATGACC 4074  
 QY 2027 AGTTCGAACCCGATCTGCAGCGCAACATCAGCG 2059  
 DB 4073 AGCCGCTCTCGACGGTCCGCGAAGCGCGG 4041

RESULT 8  
 ABL21478  
 ID ABL21478 standard; DNA; 10757 BP.  
 XX  
 AC ABL21478;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 15907.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 15907; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 10757 BP; 3125 A; 2268 C; 2226 G; 3138 T; 0 other;  
 Query Match 1.8%; Score 37.8; DB 23; Length 10757;  
 Best Local Similarity 54.1%; Pred. No. 7;  
 Matches 100; Conservative 0; Mismatches 82; Indels 3; Gaps 1;  
 QY 1283 AATCGCTTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1342  
 DB 6709 ATTGTGTTTTCGACGATGAACCAAGCTCCTGGATGCATCTTCGAGCGAAAGTTTCAGCGG 6768  
 QY 1343 CCATGCTCGATATGTACGCGCGGTATCCCTGCGGACGACGACGACGACGACGACGACGACG 1402  
 DB 6769 GT---GACTTTATCATCCGCGCGGTGACGCGCGGATCTTTATGTTATTGAATCGT 6825  
 QY 1403 ATGTAATCGCAAGAAATTCAAAGCGGACGACGACGACGACGACGACGACGACGACGACGACG 1462  
 DB 6826 AAGTAAACAAACAAATTTCAATCAAACTCCATGATTAATGACACACTTCTGTTTGCCA 6885  
 QY 1463 AGAGT 1467  
 DB 6886 AAGT 6890  
 RESULT 9  
 ABL18652/c  
 ID ABL18652 standard; DNA; 9845 BP.  
 XX  
 AC ABL18652;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7429.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 7429; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and



SQ Sequence 1077 BP; 338 A; 228 C; 225 G; 286 T; 0 other;  
 Query Match 1.7%; Score 37; DB 22; Length 1077;  
 Best Local Similarity 52.2%; Pred. No. 3.6;  
 Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
 QY 1377 CGACAAGTCCCGATATATTCAAGAATGTATACGACAAAGAAATTTCAAGGCGACACGAA 1436  
 DB 135 CGTTAAGCTCAAGGACATTTTGAACCTCCAGATTCAACAAGAGAGAGAAAACCTTCAAGAAGCT 194  
 QY 1437 GAAGTATCGACTCGTATTGCGACAAGAGTGTGGTCTTATAGCGACAAGTTCCTATGC 1496  
 DB 195 TCTGAGTCTGACTTGATTCCATCAAGGATTTGACTTCTTCTAACTACTGTTGTAAGGA 254  
 QY 1497 CATGCTCAAGTCCATGGACAAGGAAAAGCTTTGCCAAG 1533  
 DB 255 CCATACAAAGTCTCGACACAGGAGAGACAGACAAAG 291  
 RESULT 12  
 AAH01671/c  
 ID AAH01671 standard; DNA; 810 BP.  
 XX  
 AC AAH01671;  
 XX  
 DT 24-JUL-2001 (first entry)  
 XX  
 DE Fonseca pedrosol nucleotide sequence SEQ ID NO:1664.  
 XX  
 KW Species specific; genus specific; family specific; probe; detection;  
 KW identification; algal; archaeal; bacterial; fungal; parasitological;  
 KW microorganism; diagnosis; translation elongation factor Tu; toxin;  
 KW translation elongation factor G; RecA recombinase; resistance;  
 KW catalytic subunit of proton-translocating ATPase; antimicrobial;  
 KW vaccine; primer; ds.  
 XX  
 OS Fonseca pedrosol.  
 XX  
 PN WO200123604-A2.  
 XX  
 PD 05-APR-2001.  
 XX  
 PF 28-SEP-2000; 2000MO-CA01150.  
 XX  
 PR 28-SEP-1999; 99CA-2283458.  
 PR 19-MAY-2000; 2000CA-2307010.  
 XX  
 PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.  
 XX  
 PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;  
 PI Picard FJ, Roy PH;  
 XX  
 DR WPI; 2001-245006/25.  
 XX  
 PT Nucleic acid sequences are used to generate universal probes and  
 PT primers which can be used to identify and detect the presence of algal,  
 PT archaeal, bacterial, fungal and parasitological species in a test sample -  
 XX  
 PS Claim 24; Page 1264-1265; 1580pp; English.  
 CC  
 CC The present invention describes a method for generating a repository of  
 CC nucleic acids of tuf, fus, atpD and/or recA genes from which probes  
 CC and/or primers are derived. The method comprises amplifying the nucleic  
 CC acids of determined algal, archaeal, bacterial, fungal and parasitological  
 CC species with a combination of defined primer pairs. The method can be  
 CC used for producing probes and/or primers for detecting one or more  
 CC related microorganisms e.g. algal, archaea, bacteria, fungi and  
 CC parasites, for universal detection and for specific and ubiquitous  
 CC detection and identification of an algal, archaeal, bacterial, fungal  
 CC and parasitological species, genus, family and group. A nucleic acid (I)  
 CC obtained using the method of the invention can be used for the universal  
 CC detection of any bacterium, fungus or parasite in a sample and for the  
 CC detection of at least one antimicrobial agent resistance gene or at

CC least one toxin gene. hexA nucleic acids are used for the specific and  
 CC ubiquitous detection and for identification of Streptococcus pneumoniae.  
 CC (I) can be used to design a therapeutic agent which is effective against  
 CC microorganisms. Microbial species or genus or family or phylum or group  
 CC which can be detected include Abiotrophia adiacens, Bordetella sp.,  
 CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,  
 CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,  
 CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests  
 CC provides faster results than substrate specificity tests as results can  
 CC be determined in an hour and improved accuracy is also achieved.  
 CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes  
 CC which are given in the exemplification of the present invention.  
 XX  
 SQ Sequence 810 BP; 167 A; 254 C; 228 G; 161 T; 0 other;  
 Query Match 1.7%; Score 36.8; DB 22; Length 810;  
 Best Local Similarity 46.2%; Pred. No. 3.6;  
 Matches 122; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
 QY 192 TGGCGTGGTTATCTTCGCTGGCGGATGTACCGGTATACAGTGTCCGATCAGGGCTCAT 251  
 DB 305 TGGCGATGTACGCTCGGGGTGGGATCCAGGTGTCCAGCTGCTCGAGGAGCTCGTCAA 246  
 QY 252 CTTTACCACCACTGCGGATACGGTGTCTATCCAGAGCAAGCAAGCGGTGGATCAGCA 311  
 DB 245 TCTTCTCGACACCAATGTCCGGCTCGCGGCTCAATGGCGCAAAAGGCGGACCCATGA 186  
 QY 312 CTATCTCGCGCATGGTTTCGTTTCTCGCAGCATGGGTGAGAGCTTCGGATTCGGGTCT 371  
 DB 185 CGATGGGAGTGTCTGTCACCCCTCGAAGCGGTAGCTGGAGAGAGCTCAGCATCTCCATCT 126  
 QY 372 TTCCGTGAAGTATCTGCGCAAGATGCTGAAGGTAAGGACAAGGTAGAGGACAGCTCAA 431  
 DB 125 CGACGAGTCCAAATCTCTTGTCTCAATGGCATCGACCTTGTGACGAAGAGGACAA 66  
 QY 432 GGGTATCACTGACGAGATGGAGCG 455  
 DB 65 TGGCGTTGACACCGACCTGGCGGG 42  
 RESULT 13  
 AAF86431/c  
 ID AAF86431 standard; DNA; 349980 BP.  
 XX  
 AC AAF86431;  
 XX  
 DT 29-OCT-2001 (first entry)  
 XX  
 DE Pyrococcus abyssi genomic fragment #1.  
 XX  
 KW Hyperthermophilic archaeon; hyperthermophilic protein; ds.  
 XX  
 OS Pyrococcus abyssi.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 300001..349980  
 FT /\*tag= a  
 FT /note= "This sequence overlaps with the 5' end of  
 FT AAH41223"  
 XX  
 PN FR2792651-A1.  
 XX  
 PD 27-OCT-2000.  
 XX  
 PF 21-APR-1999; 99FR-0005034.  
 XX  
 PR 21-APR-1999; 99FR-0005034.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.  
 XX  
 PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;  
 PI Querellou J, Weissbach J, Saurin W, Heilig R;







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 02:43:56 ; Search time 62 Seconds  
(without alignments)  
10580.348 Million cell updates/sec

Title: US-10-008-355-1  
Perfect score: 2139  
Sequence: 1 atcgaatgaattaaaag.....aagagctgaagtgtatctaa 2139

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1717.4	80.3	1974	4	US-09-221-017B-726
2	207.4	9.7	1317	4	US-09-221-017B-382
3	164.4	7.7	561	4	US-09-221-017B-12
4	153.8	7.2	2384	4	US-09-221-017B-1045
5	36.4	1.7	1090	4	US-09-289-349-7
6	36.2	1.7	2280	1	US-08-415-818-5
7	36.2	1.7	2280	2	US-08-894-236-5
8	36.2	1.7	2280	5	PCT-US96-01444-5
9	36	1.7	7218	1	US-08-232-463-14
10	35.2	1.6	449	2	US-08-825-782-2
11	34	1.6	5837	1	US-07-686-340-1
12	34	1.6	5837	1	US-08-004-139B-1
13	34	1.6	5837	1	US-08-117-491-1
14	34	1.6	5837	2	US-08-271-364A-1
15	34	1.6	5837	2	US-08-811-492-1
16	34	1.6	5837	2	US-08-222-715B-1
17	34	1.6	5837	5	PCT-US96-10545A-1
18	33.6	1.6	23673	4	US-09-773-816-1
19	33.6	1.6	50937	4	US-09-428-517-1
20	33.2	1.6	939	4	US-09-105-390-45
21	33.2	1.6	1020	4	US-09-105-390-61
22	33.2	1.6	2612	4	US-09-105-390-7
23	32.4	1.5	710	4	US-08-998-416-603
24	32.2	1.5	3468	1	US-07-951-715A-2
25	32.2	1.5	3468	2	US-08-459-448A-2
26	32.2	1.5	3468	3	US-08-459-595A-2
27	32.2	1.5	3468	3	US-08-459-504B-2

28	32.2	1.5	3468	3	US-08-459-444-2	Sequence 2, Appli
29	32.2	1.5	3468	3	US-09-053-549-3	Sequence 3, Appli
30	32.2	1.5	3468	4	US-09-547-422-2	Sequence 2, Appli
31	32	1.5	289	4	US-09-007-005-17	Sequence 17, Appli
32	32	1.5	289	4	US-09-244-796-17	Sequence 17, Appli
33	32	1.5	823	4	US-08-998-416-551	Sequence 551, App
34	32	1.5	1947	1	US-07-951-715A-3	Sequence 3, Appli
35	32	1.5	1947	2	US-08-459-448A-3	Sequence 3, Appli
36	32	1.5	1947	3	US-08-459-595A-3	Sequence 3, Appli
37	32	1.5	1947	3	US-08-459-504B-3	Sequence 3, Appli
38	32	1.5	1947	3	US-08-459-444-3	Sequence 3, Appli
39	32	1.5	1947	4	US-09-547-422-3	Sequence 3, Appli
40	32	1.5	2845	1	US-08-289-653-1	Sequence 3, Appli
41	32	1.5	3468	1	US-07-951-715A-4	Sequence 4, Appli
42	32	1.5	3468	1	US-07-951-715A-8	Sequence 8, Appli
43	32	1.5	3468	2	US-08-459-448A-4	Sequence 8, Appli
44	32	1.5	3468	2	US-08-459-448A-8	Sequence 8, Appli
45	32	1.5	3468	3	US-08-459-595A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-221-017B-726  
; Sequence 726, Application US/09221017B  
; Patent No. 644799  
; GENERAL INFORMATION:  
; APPLICANT: Ross, Bruce C.  
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
; NUMBER OF SEQUENCES: 1120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows  
; SOFTWARE: FASTSEQ For Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/221.017B  
; FILING DATE: 23-DEC-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PPI182  
; FILING DATE: 31-DEC-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PPI546  
; FILING DATE: 30-JAN-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PP2911  
; FILING DATE: 09-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU98/01023  
; FILING DATE: 10-DEC-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Monroy, Gladys H  
; REGISTRATION NUMBER: 32,430  
; REFERENCE/DOCKET NUMBER: 27340-20021.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-813-5600  
; TELEFAX: 650-494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 726:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1974 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: Double  
; TOPOLOGY: circular

```

: MOLECULE TYPE: DNA (genomic)
:
: HYPOTHETICAL: NO
:
: ANTI-SENSE: UNKNOWN
:
: ORIGINAL SOURCE:
:
: ORGANISM: PORPHYROMONAS GINGIVALIS
:
: FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: 1..1974
:

```

Query Match	80.3%;	Score 1717.4;	DB 4;	Length 1974;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1730;	Conservative	0; Mismatches	1; Indels	2; Gaps
Qy	409	GACAAGGTAGAAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAAGCT	468	
Db				
Qy	1	GACAGGTAGAGGACAGCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAAGCT	60	
Db				
Qy	469	CAGGAGGTATGCCAAGAAGCTGGCCAAAAAAGAAATGCAGACGAGAACCAACTCTGCATC	528	
Db	61	CAGGAGGTATGCCAAGAAGCTGGCCAAAAAAGAAATGCAGACGAGAACCAACTCTGCATC	120	
Qy	529	GTAGAGCCTTTCTATTCCAAACAGCAATACTTCCCTCATCTGCTTACGATGTATTCAAGGAC	588	
Db	121	GTAGAGCCTTTCTATTCCAAACAGCAATACTTCCCTCATCTGCTTACGATGTATTCAAGGAC	180	
Qy	589	GTTCGTATGTATTTCCTCCTCCAGCTCTGTAGTAAAGTTTCGGAGGGGATACGGACAAC	648	
Db	181	GTTCGTATGTATTTCCTCCTCCAGCTCTGTAGTAAAGTTTCGGAGGGGATACGGACAAC	240	
Qy	649	TGGATGTGGCGCGGTCCACAGCGCGCACTTCACGCGTATTCGCGGTATGCGCGGTGCCGAC	708	
Db	241	TGGATGTGGCGCGGTCCACAGCGCGCACTTCACGCGTATTCGCGGTATGCGCGGTGCCGAC	300	
Qy	709	AACGGCGCGGCGCAATACAGCAAGGACAAATAAACCCCTATAAGCCGCTTACTTCGCTGCC	768	
Db	301	AACGGCGCGGCGCAATACAGCAAGGACAAATAAACCCCTATAAGCCGCTTACTTCGCTGCC	360	
Qy	769	GTATCCATGCAAGGCTACAAGGCTGACGACTATGCATGACCATCGGTTTCCCGGGCAGT	828	
Db	361	GTATCCATGCAAGGCTACAAGGCTGACGACTATGCATGACCATCGGTTTCCCGGGCAGT	420	
Qy	829	ACGSGATCGCTACCTCACTCTTTGGGGTGTGGAAGATCGTATCGAAACAGAGAACCAATCCT	888	
Db	421	ACGSGATCGCTACCTCACTCTTTGGGGTGTGGAAGATCGTATCGAAACAGAGAACCAATCCT	480	
Qy	889	CGTATCGAAGTTTCGGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAG	948	
Db	481	CGTATCGAAGTTTCGGGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAG	540	
Qy	949	GCTACCCGTATCAAAATATCCAGCAAGTATGCTCAGAGTGCCTAACTATTGGAAGAAATTCG	1008	
Db	541	GCTACCCGTATCAAAATATCCAGCAAGTATGCTCAGAGTGCCTAACTATTGGAAGAAATTCG	600	
Qy	1009	ATCGGTATGAACCCGCTCTCGCTCGCTTTGACGTGATAGGTCGTAAAGCGTCCGCGAGGAA	1068	
Db	601	ATCGGTATGAACCCGCTCTCGCTCGCTTTGACGTGATAGGTCGTAAAGCGTCCGCGAGGAA	660	
Qy	1069	AGAGCATTCGACACTGGATCCGTAAGAACGGCAAGAGTGTCTTATGGCGATGTATTG	1128	
Db	661	AGAGCATTCGACACTGGATCCGTAAGAACGGCAAGAGTGTCTTATGGCGATGTATTG	720	
Qy	1129	TCCTCTCTGAAAAGGCTTATAGGAGAGGACCAAGGCCAACCGTGAGATGACTATTATG	1188	
Db	721	TCCTCTCTGAAAAGGCTTATAGGAGAGGACCAAGGCCAACCGTGAGATGACTATTATG	780	
Qy	1189	AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTTCGTTTTTGACAGTTTGCACAGCATTG	1248	
Db	781	AGCGAGACGCTCTTCGGTGGTACCGAGGTGGTTTCGTTTTTGACAGTTTGCACAGCATTG	840	
Qy	1249	GCTACAAATCCTGATGCTATCGCGGTATCCCTCAATCGCTTTCGACGACAAAGTACAAAGAC	1308	
Db	841	GCTACAAATCCTGATGCTATCGCGGTATCCCTCAATCGCTTTCGACGACAAAGTACAAAGAC	900	

Qy	1309	TAC	TCCCTCCCTCGCTCGACCGTGAAGTGTGCGCCGCCATGCTCGATATTGTATACGCCGCGCT	1369
Db	901			960
		TAC	TCCCTCCCTCGCTCGACCGTGAAGTGTGCGCCGCCATGCTCGATATTGTATACGCCGCGCT	
Qy	1369	ATCC	TCCGCGACAAAGCTCCCGGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC	1428
Db	961			1020
		ATCC	TCCGCGACAAAGCTCCCGGATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGC	
Qy	1429	GACACG	AAAGATATCGACACTTCGCTATTTCGACAAAGAGTGGTTCCTTTATAGCGACAAG	1488
Db	1021			1080
		GACACG	AAAGATATCGACACTTCGCTATTTCGACAAAGAGTGGTTCCTTTATAGCGACAAG	
Qy	1489	TTCCAT	GCCATGCTCAAGTCCATGNCACAAAGGAAAAGTTTGCAAGGCTATCGCAGAAAGAT	1548
Db	1081			1140
		TTCCAT	GCCATGCTCAAGTCCATGNCACAAAGGAAAAGTTTGCAAGGCTATCGCAGAAAGAT	
Qy	1549	CCGCG	ACTAGAGCTTTCCAAAGACGCTAAATAGCTGCTGCTCGCGCTATTTCAGCGCCGATGCG	1608
Db	1141			1200
		CCGCG	ACTAGAGCTTTCCAAAGACGCTAAATAGCTGCTGCTCGCGCTATTTCAGCGCCGATGCG	
Qy	1609	ATGCC	AAATGCCCTATGCCATTGAGAAAGGCGCAAGCGCTCTTTTCTTTTCCGGTTCGCGTGAAG	1668
Db	1201			1260
		ATGCC	AAATGCCCTATGCCATTGAGAAAGGCGCAAGCGCTCTTTTCTTTTCCGGTTCGCGTGAAG	
Qy	1669	ATGT	TACCCCGGAGCTGCTGCGCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGCGC	1728
Db	1261			1320
		ATGT	TACCCCGGAGCTGCTGCGCGAGCGATGCCAACTTCACCATGCGTATGAGCTACGCGC	
Qy	1729	TCCAT	CAAGGGATATGAACCGCAGGACGCTGCTGGTACAACATATCATACGACAGGCAAG	1788
Db	1321			1380
		TCCAT	CAAGGGATATGAACCGCAGGACGCTGCTGGTACAACATATCATACGACAGGCAAG	
Qy	1789	GGCG	TATTGGAGAAAGCAGGAT--CCTTAAGAGCGATGAGTTTCCCGTACAGAGAAATATCC	1846
Db	1381			1440
		GGCG	TATTGGAGAAAGCAGGAT--CCTTAAGAGCGATGAGTTTCCCGTACAGAGAAATATCC	
Qy	1847	TCGAC	CTTCCGACCAAAACATATGGTTCGCTATGCCGAGAACGGTCAGCTCCATATCG	1906
Db	1441			1500
		TCGAC	CTTCCGACCAAAACATATGGTTCGCTATGCCGAGAACGGTCAGCTCCATATCG	
Qy	1907	CTTTCC	TATCGAACACGACATCACGGGCGGTAACTTCGGGTAGCCCCGTATTCGATAGA	1966
Db	1501			1560
		CTTTCC	TATCGAACACGACATCACGGGCGGTAACTTCGGGTAGCCCCGTATTCGATAGA	
Qy	1967	ACGG	CGCTCTGATCGGTCTGCTTTTCGATGGCACTGCGCAAGCTATGAGTGGTACATCG	2026
Db	1561			1620
		ACGG	CGCTCTGATCGGTCTGCTTTTCGATGGCACTGCGCAAGCTATGAGTGGTACATCG	
Qy	2027	AGTT	TCGAACCGCATCTCGACGCGACAATCAGGCTGGACATCCGCTACGCTTCTTTCATGA	2086
Db	1621			1680
		AGTT	TCGAACCGCATCTCGACGCGACAATCAGGCTGGACATCCGCTACGCTTCTTTCATGA	
Qy	2087	TTGACA	AAATGGGTGCTAGTGGCCCGGCTCATCCACAGAGCTGAAGTTGATCTAA	2139
Db	1681			1733
		TTGACA	AAATGGGTGCTAGTGGCCCGGCTCATCCACAGAGCTGAAGTTGATCTAA	

```

RESULT 2
US-09-221-017B-382
; Sequence 382, Application US/09221017B
; Patent No. 644799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018

```

Qy		459	GCGCAAGCTCAGAGGTATGCCAAGAACCTGGCCAAAAAAGAAATGCAAGACGAGAACA	518
Db		366	GACTACTCTCAAAGTTGGCTGACAAGAAGCGCGGCAAGAACTTTCTTCGCAAGAATCC	425
Qy		519	ACTCTGCATCGTAGAG-----CCTTCTATTATCCACAAGATACTCTCATCGTCTA	572
Db		426	GGGCTTTCCGTAGAGATCAAAGCCTTCTATGGGGCAATCTCTACTGATGTTTTACCAG	485
Qy		573	CGATGTATTCAAGGACCTTCGTATGTTATTTGCTCTCCAGCTCTGTAGTAAGTTCGG	632
Db		486	AAGACTTATACGGATGTTCCGACTGGTGAGGAGCACCTCCCACCAGCATTTGGCAATTCGG	545
Qy		633	AGGGATACGGAACAATGGATGTGGCGCGCTACACGGGGCACTTCAGCGTATTCGCGGT	692
Db		546	TGCCGATACGGAACAATGGATGTGGCTCGCTACTATGCGCACTTCTCCACTTTCGGTAT	605
Qy		693	GTATGCGCGTGCCGACAACCGCGCGCGAATACAGCAGGACAATAAACCCCTATAAGCC	752
Db		606	CTATGCGGACAAGNATGGCAATCCCGACCATACTGTAAGATANTGTTCCGGCTCANGCC	665
Qy		753	CGTTTACTTCGCTGCCGTATCCATGCAAGGCTACAAGGCTGAGCGACTATGCCATGACCAT	812
Db		666	GAAGCGTTTCTTCAATATCTCCCTTGGTGGAGTACAAGAACGACTACGCCATGATAAT	725
Qy		813	CGGTTTCCGGCGAGTACGATCGCTACCTCACTCTTCTTGGGGTGTGGAAGATCGTATCGA	872
Db		726	GGGTTTCCCCGGTACTACGACACGGCTATTTTCAGGGCTTCCGAAGTAGACGAATGGAAAAG	785
Qy		873	AAACGAGAACAATCCTCGTATCGAAGTTTCCGGGTATCAACAGGCGATCTGGAAGGAGCC	932
Db		786	CATCGACAACGATATTTCGATATCGCATCGCATGCGTGTATTCGTCAGGGTGTATGCTCAGGGA	845
Qy		933	CATGAGCGCAGATCAGGCTACCCGATCAAAATGTCAGCAAGTATGCTCAGAGTCTCTAA	992
Db		846	AATGCTGGCGGATCCTCAGATCAAAATCATGTATTGAGCTAAATATGCCCTTCGCGAGAA	905
Qy		993	CTATTGGAAGAAATTCGATTCGGTATGAACCGCGGTCTC	1029
Db		906	TGCTTTACAAACGTGCTATAGGTGCCAACTGGCGGATC	942

RESULT 3  
US-09-221-017B-12  
Sequence 12, Application US/09221017B  
Patent No. 644799  
GENERAL INFORMATION:  
APPLICANT: ROSS, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FASTSEQ for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911

; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; PRIOR APPLICATION DATA: P1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA: P11546
; APPLICATION NUMBER: P11546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA: P2911
; APPLICATION NUMBER: P2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA: PCT/AU98/01023
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc\_feature
; LOCATION: 1...561
; US-09-221-017B-12
;
Query Match 7.7%; Score 164.4; DB 4; Length 561;
Best Local Similarity 95.7%; Pred. No. 1.7e-42;
Matches 180; Conservative 0; Mismatches 6; Indels 2; Gaps 1;
;
Qy 1 ATGCAATGAAATTAAGATTTCTTCGGAGCAGCCCTGCTGTGGGTGCTTACAGG 60
Db 366 ATGCAATGAAATTAAGATTTCTTCGGAGCAGCCCTGCTGTGGGTGCTTACAGG 425
;
Qy 61 GTAGCCAAAGCCGCAAGAGCATGTGCTCTCTCAACGAATCAATCAGGAGATCTGGAT 120
Db 426 GTAGCCAAAGCCGCAAGAGCATGTGCTCTCTCAACGAATCAATCAGGAGATCTGGAT 485
;
Qy 121 CGAATGGTGTGCTGCTTACCTCCCGTGGATTCGCTCTACAGTTTCGACAAGCCG 180
Db 486 CGAATGGTGTGCTGCTTACCTCCCGTGGATTCGCTCTACAGTTTCGACAAGCCG 543
;
Qy 181 TCCATTGC 188
Db 544 TCCATGCG 551
;
RESULT 4
US-09-221-017B-1045
; Sequence 1045, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA: P11546
; APPLICATION NUMBER: P11546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA: P2911
; APPLICATION NUMBER: P2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA: PCT/AU98/01023
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 1045:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2384 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc\_feature
; LOCATION: 1...2384
; US-09-221-017B-1045
;
Query Match 7.2%; Score 153.8; DB 4; Length 2384;
Best Local Similarity 55.0%; Pred. No. 1e-38;
Matches 324; Conservative 0; Mismatches 262; Indels 3; Gaps 1;
;
Qy 1544 AAGATCCGGCAGTAGAGCTTTCCNAGAGCGTAATAGTGTGCTCGCGCTATTTCAGGCCG 1603
Db 31 AAGATCCATATGGTACTCTTCGCGCTCTTCGGTGTTCGACGAATACCGCAAACTCTACAAACG 90
;
Qy 1604 ATGCGATGGCCCAATGCTTATGCCATTGAGAAGGCAAGCGTCTTTCTTGGCGGTTTGC 1663
Db 91 AACTTCGCTCCCTATGACGATCCCTATTCTAAGGCGCACACGCCACCTACATTGCGAGTCTCT 150
;
Qy 1664 GTGAGATGTACCCCGGACGTGCTCTGCCGAGCGATGCCAATTCACCATGCGTATGAGCT 1723
Db 151 TGAATATGGATGGCGATCAGGATCAATTCGCGGATGCTAACCTGACACTTCGTTTCACT 210
;
Qy 1724 ACGCTCCATCAAGGGATATGAACCGCAGGACGCTGCTGTACACATATCATACGACAG 1783
Db 211 ATGCTCAAGTGAAGGGTATTCACCCCGTGCACAAATGTTTACTAGCGACATCAAAACCAT 270
;
Qy 1784 GCAAGGCGTATTGGAGAAGCAGGATCCTAAGACGCGATAGTTTGGCGGTACAGGAGAATA 1843
Db 271 TGGATGTGTGATGGAAAAAGAGATCCCGATTAATTCGGATTTGTAGTCTGATCCCAAGC 330
;
Qy 1844 TCCTCGACCTCTTCCCGCACCACCAAACTATGTCGCTATGCCGA---GAACGGTCAAGCTCC 1900
Db 331 TGAAGCCGTATACGAGCGTAAAGACTTCGCGCGTATTCGCGATCGCAGCGGTCGCTGTC 390
;
Qy 1901 ATATCGCTTTTCCCTATCGAACAACGACATCAGGCGCGTAACTCGCGTAGCCCGCTATTTCG 1960
Db 391 CTGTAGCTTTTTCGCCCAACACACATACAAACCGCGCAACTCAGGACGCTCCGGTCTATGA 450
;
Qy 1961 ATAAGAAGCGCGCTCTGATCGGCTCTTCTTCGATGGCAAGCTATGATGAGTGGT 2020
Db 451 ATGCCAAGCGGGAAGTATGCGGTCTCACTCACTGAGTGGAGGAGTGGTGGCG 510
;
Qy 2021 ACATCGAGTTTGAACCCGATCTGACGCGCAATACGCGTGGACATCCGCTACGCTTCTCT 2080

RESULT 7  
US-08-894-236-5  
; Sequence 5, Application US/08894236  
; Patent No. 5939263  
; GENERAL INFORMATION:  
; APPLICANT: Cascieri, Margaret A.  
; APPLICANT: Linemeyer, David L.  
; APPLICANT: MacNeill, Douglas J.  
; APPLICANT: Shiao, Lin-Lin  
; APPLICANT: Strader, Catherine D.  
; APPLICANT: Tan, Carina P.  
; APPLICANT: Weinberg, David H.  
; TITLE OF INVENTION: NEUROPEPTIDE Y RECEPTOR  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Mary A. Appollina  
; STREET: P. O. Box 2000, 126 E. Lincoln Ave  
; CITY: Rahway  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07065  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/894,236  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/383,746  
FILING DATE: 03-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/415,818  
FILING DATE: 03-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Appollina Mary A.  
REGISTRATION NUMBER: 34,087  
REFERENCE/DOCKET NUMBER: 19390Y  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-3462  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2280 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 822..1937  
US-08-894-236-5

Query Match 1.7%; Score 36.2; DB 2; Length 2280;  
Best Local Similarity 55.0%; Pred. No. 0.44;  
Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 409 GACAGGTTAGAGGACACTCAAGGTATCACTGACGAGATGGAGCGTCTGGCAAAGCT 468  
Db 501 GACAGGAGGATCAGAGCGTTAAGGACATCATTTGTACATAGTAGTGTTGAGGAAAGCT 560

Qy 469 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAATGCGAGAGCAACCAACTCTGCATC 528  
Db 561 GAGGTTACATGGAACCTCTCTCTCTCAAAACAAACAAACAAACAAACCTTCTAC 620

Qy 529 GTAGAGCCT 537  
Db 621 TAATATTCT 629

RESULT 8  
PCT-US96-01444-5  
Sequence 5, Application PC/TUS9601444  
GENERAL INFORMATION:  
APPLICANT: Cascieri, Margaret A.  
APPLICANT: Linemeyer, David L.  
APPLICANT: MacNeil, Douglas J.  
APPLICANT: Shiao, Lin-Lin  
APPLICANT: Strader, Catherine D.  
APPLICANT: Tan, Carina P.  
APPLICANT: Weinberg, David H.  
TITLE OF INVENTION: NEUROPEPTIDE Y RECEPTOR  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mary A. Appollina  
STREET: P.O. Box 2000, 126 E. Lincoln Ave.  
CITY: Rahway  
STATE: NJ  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01444  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/383,746  
FILING DATE: 03-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/415,818  
FILING DATE: 03-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Appollina Mary A.  
REGISTRATION NUMBER: 34,087  
REFERENCE/DOCKET NUMBER: 19390Y  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-3462  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2280 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 822..1937  
PCT-US96-01444-5

Query Match 1.7%; Score 36.2; DB 5; Length 2280;  
Best Local Similarity 55.0%; Pred. No. 0.44;  
Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 409 GACAGGTTAGAGGACACTCAAGGTATCACTGACGAGATGGAGCGTCTGGCAAAGCT 468  
Db 501 GACAGGAGGATCAGAGCGTTAAGGACATCATTTGTACATAGTAGTGTTGAGGAAAGCT 560

Qy 469 CAGGAGGTATGCCAAGAACTGGCCAAAAAAGAAATGCGAGAGCAACCAACTCTGCATC 528  
Db 561 GAGGTTACATGGAACCTCTCTCTCTCAAAACAAACAAACAAACAAACCTTCTAC 620

Qy 529 GTAGAGCCT 537  
Db 621 TAATATTCT 629

RESULT 9  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS



SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232.463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935.313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgt-Fls  
US-08-232-463-14

Query Match 1.7%; Score 36; DB 1; Length 7218;  
Best Local Similarity 3.2%; Pred. No. 1.1;  
Matches 12; Conservative 204; Mismatches 164; Indels 0; Gaps 0;

QY 851 GGCGTGTGAAGATCGTATCGAAGCAAGCAAACTCTGTCGATCGACGGTGATCA 910  
Db 1423 RR 1364  
QY 911 AGCAAGCATCTGGAAGAGCAGCCATGACGCAGATCAGCGTATCAAATGCCA 970  
Db 1363 RR 1304  
QY 971 GCAAGTAGCTCAGAGTGCTACTATGGAACAATTCGATCGGTATGACCGCGTCTCG 1030  
Db 1303 RR 1244  
QY 1031 CTCGCTTTCAGTGATAGTCTTAAGCGTCGCGAGAAGACATTCGACACTGCATCC 1090  
Db 1243 RR 1184  
QY 1091 GTAAGACGCAGAGTGTCTGTATGCGCATGTATGTCTTCGAAAAGGCTTATA 1150  
Db 1183 RR 1124  
QY 1151 AGAAGAGCCAGGCAACCGTGAGATGACATTTATGAGCGAGACGCTCTTCGGTGTA 1210  
Db 1123 RR 1064  
QY 1211 CCAGGTGGTTCGTTTTGCA 1230  
Db 1063 GCAAGCTCCCTCGACCTGCA 1044

RESULT 10  
US-08-825-782-2  
Sequence 2, Application US/08825782  
Patent No. 5834239  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: COFACTOR A-LIKE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/825,782  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0263 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 449 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT08  
CLONE: 1654275  
US-08-825-782-2

Query Match 1.6%; Score 35.2; DB 2; Length 449;  
Best Local Similarity 53.7%; Pred. No. 0.32;  
Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0

QY 422 GACAGCTCAAGGTTATCATGACGAGATGAGCGTCTGGGCAAGCTCAGAGAGGTATGCC 481  
Db 89 GACAGATCAAGATCAAGACCGCGTGTGTAAGCGTTGGTCAAAGAAAAAGTATGATG 148  
QY 482 AGAACCTGGCCAAAAAGAAATGCACAGAGAACCACTCTGCATCGTAGACCCTTCT 541  
Db 149 AAAAGAGGCAAAACACAAAGAAAGATTGAAAAAATGAGAGCTGAAGACGGTAAA 208  
QY 542 ATTCACCAACGAATA 557  
Db 209 ATTACGACATTAAAA 224

RESULT 11  
US-07-686-340-1/C  
Sequence 1, Application US/07686340  
Patent No. 5322785  
GENERAL INFORMATION:  
APPLICANT: Comb, Donald G.  
APPLICANT: Perler, Francine  
APPLICANT: Kucera, Rebecca  
APPLICANT: Jack, William E.  
TITLE OF INVENTION: Purified Thermostable DNA Polymerase  
TITLE OF INVENTION: Obtainable From Thermococcus Litoralis  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS  
STREET: & CUSHMAN  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

US-07-686-340-1

Db 3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTATTTCACAAATCTA 3581

TELECOMMUNICATION INFORMATION:

```

: FILING DATE: 17-APR-1991
:
: PRIOR APPLICATION NUMBER: US 07/626,057
:
: FILING DATE: 11-DEC-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 07/513,994
:
: FILING DATE: 26-APR-1990
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Resnick, David S.
:
: REGISTRATION NUMBER: 34,235
:
: REFERENCE/DOCKET NUMBER: 39296C3FWC2
:
: TELECOMMUNICATION INFORMATION:
:

```

TELEPHONE: (617) 523-3400  
TELEFAX: (617) 523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5837 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-117-491-1

Query Match 1.6%; Score 34; DB 1; Length 5837;  
Best Local Similarity 57.5%; Pred. No. 4.1;  
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
Qy 1405 GTAATCGACAAGAAATTCAAAGGCGACACGAGTATGCGAGACTTCGTTATTCGACAAG 1464  
Db 3686 GTCATCGTCCAGAGTTAGTGCTTCAACACCTTCGAGATGCGAGTATTCGTTTCCCAAT 3627  
Qy 1465 AGTGTGGTTCCTTATAGGACAAGTTCATGCCATGCCATGCCTCAAGTCCA 1510  
Db 3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAAATCTA 3581

RESULT 14  
US-08-271-364A-1/c  
Sequence 1, Application US/08271364A  
Patent No. 5756334  
GENERAL INFORMATION:  
APPLICANT: PERLER, FRANCINE B.  
APPLICANT: SOUTHWORTH, MAURICE W.  
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA POLYMERASE  
TITLE OF INVENTION: FROM ARCHAEBACTERIA  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEW ENGLAND BIOLABS, INC.  
STREET: 32 TOZER ROAD  
CITY: BEVERLY  
STATE: MASSACHUSETTS  
COUNTRY: US  
ZIP: 01915

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/271.364A  
FILING DATE: 06-JUL-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/811,421  
FILING DATE: 18-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/686,340  
FILING DATE: 17-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/626,057  
FILING DATE: 11-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/513,994  
FILING DATE: 26-APR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: WILLIAMS, GREGORY D.  
REGISTRATION NUMBER: 30901  
REFERENCE/DOCKET NUMBER: NEB-101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (508) 927-5054  
TELEFAX: (508) 927-1705  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5837 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: not relevant  
US-08-271-364A-1

Query Match 1.6%; Score 34; DB 1; Length 5837;  
Best Local Similarity 57.5%; Pred. No. 4.1;  
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
Qy 1405 GTAATCGACAAGAAATTCAAAGGCGACACGAGTATGCGAGACTTCGTTATTCGACAAG 1464  
Db 3686 GTCATCGTCCAGAGTTAGTGCTTCAACACCTTCGAGATGCGAGTATTCGTTTCCCAAT 3627  
Qy 1465 AGTGTGGTTCCTTATAGGACAAGTTCATGCCATGCCATGCCTCAAGTCCA 1510  
Db 3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAAATCTA 3581

RESULT 15  
US-08-811-492-1/c  
Sequence 1, Application US/08811492  
Patent No. 5834247  
GENERAL INFORMATION:  
APPLICANT: COMB, DONALD G.  
APPLICANT: PERLER, FRANCINE B.  
APPLICANT: JACK, WILLIAM E.  
APPLICANT: XU, MING-QUN  
APPLICANT: HODGES, ROBERT A.  
APPLICANT: NOREN, CHRISTOPHER J.  
APPLICANT: CHONG, SHAORONG S.C.  
APPLICANT: ADAM, ERIC  
APPLICANT: SOUTHWORTH, MAURICE  
TITLE OF INVENTION: MODIFIED PROTEINS, METHODS OF THEIR  
TITLE OF INVENTION: PRODUCTION AND METHODS FOR PURIFICATION OF TARGET  
NUMBER OF SEQUENCES: 155  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GREGORY D. WILLIAMS; NEW ENGLAND BIOLABS, INC.  
STREET: 32 TOZER ROAD  
CITY: BEVERLY  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 01915

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/811.492  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/580,555  
FILING DATE: 29-DEC-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/496,247  
FILING DATE: 28-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/146,885  
FILING DATE: 03-NOV-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/004,139  
FILING DATE: 09-DEC-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams, Gregory D  
REGISTRATION NUMBER: 30901  
REFERENCE/DOCKET NUMBER: NEB-036C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 508-927-5054

```
;
; TELEFAX: 509-927-1705
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
;
US-08-811-492-1

Query Match      1.6%; Score 34; DB 2; Length 5837;
Best Local Similarity 57.5%; Pred. No. 4.1;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1405 GTAATCGACAAGAAATTCAAAGGGGACACGAAAGATATGCAGACTTCGTATTGACAAG 1464
      ||||| | || | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 3686 GTCATCGTCCAGAGTTAGTGCTTCAACACCTTCGAGATGCAGTATTCTTTTCGCCAAT 3627

Qy 1465 AGTGTGGTTCCTTTATAGCGACAAGTTCCATGCCATGCTCAAGTCCA 1510
      ||| | | || | || | || | || | || | || | || | || | || | || |
Db 3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTCACAAATCTA 3581

Search completed: May 23, 2003, 04:44:48
Job time : 122 secs
```

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 03:33:01 ; Search time 529 seconds  
(without alignments)  
5339.261 Million cell updates/sec

Title: US-10-008-355-1  
Perfect score: 2139  
Sequence: 1 atgcaatgaattataaag.....aagagctgaattgatctaa 2139

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 828747 seqs, 650231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published\_Applications\_NA:\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
  - 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2139	100.0	2139	9	US-10-008-355-1
2	40.8	1.9	867	9	US-10-123-155-20
3	36.2	1.7	345	9	US-09-738-626-2081
4	36.2	1.7	2281	10	US-09-900-497-1
5	36	1.7	292	10	US-09-294-093B-856
6	36	1.7	369	10	US-09-878-574-5240
c 7	35.2	1.6	671	9	US-10-184-644-346
c 8	35.2	1.6	671	9	US-10-184-634-346
9	35	1.6	1896	9	US-10-213-990-53
10	35	1.6	2060	9	US-10-213-990-52
11	34.4	1.6	489	9	US-10-184-644-116
12	34.4	1.6	489	9	US-10-184-634-116
13	34	1.6	268	10	US-09-923-876-5908
c 14	34	1.6	50000	9	US-10-152-724A-22
15	33.6	1.6	50937	9	US-09-808-880-1
c 16	33.4	1.6	802	9	US-10-184-644-312
c 17	33.4	1.6	802	9	US-10-184-634-312
18	33.2	1.6	256	10	US-09-878-574-9837
19	33.2	1.6	607	9	US-10-123-155-344

ALIGNMENTS

RESULT 1

US-10-008-355-1

; Sequence 1, Application US/10008355

; Patent No. US20020164759A1

; GENERAL INFORMATION:

; APPLICANT: Travis, James

; APPLICANT: Potempa, Jan S

; APPLICANT: Banbula, Agnieszka

; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101

; CURRENT APPLICATION NUMBER: US/10/008.355

; CURRENT FILING DATE: 2001-11-08

; PRIOR APPLICATION NUMBER: US 60/246,827

; PRIOR FILING DATE: 2000-11-08

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 2139

; TYPE: DNA

; ORGANISM: Porphyromonas gingivalis

US-10-008-355-1

Query Match 100.0%; Score 2139; DB 9; Length 2139;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAAAATGAATTAATAAGATTTCTTCGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60

Db 1 ATGCAAAATGAATTAATAAGATTTCTTCGAGCAGCCCTGCTGTTGGGTGCTTCAGGG 60

Qy 61 GTAGCAAGCCGACAAAGGATGTGGCTCTCTCAACGAACATCAATCAGAGAAATCTGGAT 120

Db 61 GTAGCAAGCCGACAAAGGATGTGGCTCTCTCAACGAACATCAATCAGAGAAATCTGGAT 120

Qy 121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCCGCTTACAGTTTCAGAACCG 180

Db 121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTCCGCTTACAGTTTCAGAACCG 180

Qy 181 TCCATTCGAATGCGTGAGCTTTTCGCTGCGGATGTACCGGATACACATGTCCCAT 240

Db 181 TCCATTCGAATGCGTGAGCTTTTCGCTGCGGATGTACCGGATACACATGTCCCAT 240



```
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 20
; LENGTH: 867
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-20

Query Match      1.9%, Score 40.8; DB 9; Length 867;
Best Local Similarity 7.6%; Pred. No. 0.01;
Matches 55; Conservative 237; Mismatches 427; Indels 1; Gaps 3

QY   691 GTGTATCGCGTCCGACAAACGGCCGCCCAATAACAGCAGGACAATAAACCTATAAG 750
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   33 GRFDRNRNRPNIILVLTDDDELGSVMYNTRRIMEQGGAHFNAFYTPMCCPSR 92
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   751 CCCGTTACTTCGCTCGCGTATCATCAAGGCACAAAGCTGACGACTATGCCATGACC 810
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   93 SSILTKGVVHNHNYTNNECSSFSWQAQHESRTFAVYLNSTGYRTAFGKYLYEINGSY 152
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   811 ATCGGTTTCCCGGCAGTAGCGATCGCTACTCACTCTCTTGCGGTGTGGAAGATCGTATC 870
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   153 VPPGWKEWGLLKNSRFNYTLCRNGVKHGSDYSKDYLTDLITNDVSFEFTSKMYP 212
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   871 GAAACGAGACATCTCGTATCGAAGTTCGGGTATCAAGCAGGACATCTGGAAGAA 930
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   213 HRPVLMIWSHAHPGEDSAPYSLRFLPNASOHTTSPSYNPDPDKHMIRYTGPMPKI 272
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   931 GCCATGACGCGAGATCAGGCTACCGGTATCAAATATGCCACAAGTATGTCACAGTGTCT 990
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   273 MEFTNMLQRKLQTLMSVDDSMETIYNMLVETGELDNITYIVYTDHGHIGQFGLVKGS 332
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   991 AACTATTGGAAGAATTCGATCGGTATGAACCGCGCTCTCGCTCGTCTTGACGTGATAGTT 1050
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   333 MPYEFDIRVPYVRGNPVEAGCLNPHIVLNLDTAPTILDIAGLDIPADMKGSIKLKLD 392
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   1051 CGTAAGCGTCCGAGGAAGAGACATCCACACATGGATCCGTAAGACGGCAAGATGCT 1110
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   393 ERPYNRPHLKKMKWRDSFLVERKLLHKRDNDKVDAQEENFLPKYORVKDLQRAEQ 452
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   1111 GTCATGCGGATGATGTCTCTCTCGAAAAGCTTATAAGGAAGGACCAAGCCCAAC 1170
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   453 TACEQLGKWCVEDATGKLLHKCK -GPMRLGGRSALSNLVPKYYGGSEACTCDSDGY 511
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   1171 CGTGAGATGACTATTATTGAGCGAGACGCTCTCGGTGGTACCGAGGTGGTCTGTTTTGCA 1230
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   512 KLSLAGRKLFKKYKASYVRSIRSVAIEVDGRVVHVGDAAPRNLTKRHWCPAP 571
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   1231 CAGTTTCCCAACGATGGCTACAAATCCTGATGCTCATGCGGTATCCTCAAAATCGGTT 1290
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   572 EDQDDKDGDFSGTGLPLPDVSAANPIKVTHRCYILENDTVQCDDLKYSLQAWKDHLHI 631
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   1291 GACCAAGTACAAGACTACCTCCCCTCGCTCCAGCCTAGTAGTGCTGCCCGCATGCTC 1350
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   632 DHEJETLQNLIKNUREVRGHLKKRPEECDCCHKTSYHQHGRGLKHRRSSLUHPRFKGLQE 691
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY   1351 GATATTGTACGCCGCGGTATCCCTCGCCGACAAAGCTCCCCGATATATCAAGAAATGTAATC 1410
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB   692 KDKVYLLEQRKKKLRKLLRLQNNDTCSMPGLTCTFDHNOHQHOTAPFWTLGFPCACTS 751
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 3  
US-09-738-626-2081





Db 420 SSS..MMTSMWT.Y...T...SHSS..CYCYMTCYMB..HH.Y..HCSSCCSY.T. 361

Qy 1877 GCTATGCCGAGAACGCTCAGCTCCATATCGCTTTCCTATCGAACCAACGACATCATCGGCG 1936

Db 360 CYHTY.VVY.M..M..Y.YY.....YYYYT..YYT.VVYMW.S.YHB.HSHSHSS 301

Qy 1937 GTAACTCCGGTAGCCCCGATTTCGATAGAACGGCCGCTCGATCGCTGCTTG 1987

Db 300 SS..Y.Y.M.MYCY.M..M.T.MYCY..MMBSBHSHSSSSSSSSSSYTYT. 250

## RESULT 8

```

US-10-184-634-346/C
; Sequence 346, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430RIC217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-346

```

## RESULT 9

```

US-10-213-990-53
: Sequence 53, Application US/10213990
: Publication NO. US20030082595A1
: GENERAL INFORMATION:
: APPLICANT: Jiang, Bo
: APPLICANT: Bussey, Howard
: APPLICANT: Storms, Reg
: APPLICANT: Roemer, Terry
: TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL
: TITLE OF INVENTION: ENZYMES AND METHODS OF USE
: FILE REFERENCE: 10182-019-999
: CURRENT APPLICATION NUMBER: US/10/213,990
: CURRENT FILING DATE: 2002-08-05
: NUMBER OF SEQ ID NOS: 72
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 53
: LENGTH: 1896
: TYPE: DNA
: ORGANISM: Aspergillus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(1896)
US-10-213-990-53

```

## RESULT 10

```

US-10-213-990-52
? Sequence 52, Application US/10213990
? Publication No. US20030082595A1
? GENERAL INFORMATION:
? APPLICANT: Jiang, Bo
? APPLICANT: Bussey, Howard
? APPLICANT: Storms, Reg
? APPLICANT: Roemer, Terry
? TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL
? FILE OF INVENTION: ENZYMES AND METHODS OF USE
? FILE REFERENCE: 10182-019-999
? CURRENT APPLICATION NUMBER: US/10/213,990
? CURRENT FILING DATE: 2002-08-05
? NUMBER OF SEQ ID NOS: 72
? SOFTWARE: FastSEQ for Windows Version 4.0
? SEQ ID NO 52
? LENGTH: 2060
? TYPE: DNA
? ORGANISM: Aspergillus
US-10-213-990-52

```

**D**b      Qy    1937 GTAACTCCGGTAGCCCCGATTTCGTATAAGAACGCCGCTCGATCGTGTTG 1987  
               : ; : ; : | : : : : :  
               : ; : ; : | : : : : :  
**D**b      300 SS..Y.M.M.YCY.M...M.T.MYC.MMBSHSHSSSSSSSSSYTK 250

[illegible]

```

RESULT 11
US-10-184-644-116
: Sequence 116, Application US/10184644
: Publication No. US20030044930A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Jian
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3430RIC227
: CURRENT APPLICATION NUMBER: US/10/184,644
: CURRENT FILING DATE: 2002-06-28
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 116
: LENGTH: 489
: TYPE: PRT
: ORGANISM: Homo Sapien
US-10-184-644-116

```

Query Match	1.6%	Score 34.4:	DB 9;	Length 489;
Best Local Similarity	8.3%	Pred. No. 1;		
Matches	37;	Conservative 116;	Mismatches 291;	Indels 0; Gaps 0;

  

QY	66	CAAAAGCCGACAAAGCGCATGTGGCTCTCCACACGAACTCAATCAGGAGATCTGGCATCGAAT	125
DB	1	MEAPDYELSVREQLFHERIERECIIISTLLFATLYILCHIFILTRFKKPAEFTTVDDEDATV	60
QY	126	CGGTGAGTCGGCTTTACGCTCCGCTGGATCGCTCTACAGTTTCGACAAAGCCGCTCCAT	185
DB	61	NKIALELCFTTALTAGVALLPFSIISNEVLLSPRNYIQWLNGSLIHGLNWLFLFP	120
QY	186	TGCCAATGCCGTGTTATCTTCGTGGCGGATGACCGGTATCATCAGTGTCCGATCAGGG	245
DB	121	NLSLIFLMPFAFYFTSESGFAGSRKGVLRGYVTVVMLMLLTLLVLGMVWVASAIVDNK	180
QY	246	CTGATCTTTACCAACACCACCTCGGGATACGGTCTATCAGAGCCAAAGCACCGGTGA	305
DB	181	ANRESLYDFWEYIIPYLYVSCISFGLVLLLVCTPLGLARMPFSVTKLVLKPRLLEDLEQ	240
QY	306	TCACGAGCTATCTCGCGCATGCTTTCGTTCTTCGACAGATGGGTGAGGAGCTTCGATGCC	365
DB	241	LYCSAFEAALTRICNPTSCWLPDLMELLHROVLALQTVLLEKRRKASAWQRNLGYP	300
QY	366	GGGTCTTCCGTGAAGTATCTCGCAAGATCGTGAAGGTAAACGGACAAAGGTAGAGGACA	425
DB	301	LAMLCLLVLITGLSVLIVAIHLELLIDEAAPRGMOGTSLGOVFSFKISGFEAVTIOVLLI	360

```

Qy 426 GCTCAAGGGTATCACTGACGAGATGGAGCGTCTGCGCAAAGCTCAGGAGGTATGCCAAGA 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 FYLMVSSVGFYSPLFRSRPRRHDTAMTQIIIGNCVCLLVSSALPVSRTLGLTRFDL 420
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 486 ACTGCGCCAAAAGAAAGAAATGCAGA 509
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 421 LGDFGRENWLGNFYIVFLYNAFA 444
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-10-184-634-116
; Sequence 116: Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430RIC217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 116
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-184-634-116

```

Query Match	1.6%	Score 34.4	DB 9	Length 489
Best Local Similarity	8.3%	Pred. No. 1		
Matches	37	Conservative 116	Mismatches 291	Indels 0
Gaps				
Qy	66	CAAGCCGACAAAGCGATGGCTCTCAACGAACTCAATCAGGAGAACTGGATCGAAT	125	
Db	1	MEAPDYEVLSRQSLFHERIRECIISLTLFATLYILCHIFLTRFKKPAEFTTVDDEDATV	60	
Qy	126	CGGTGAGCTGGGTTTACGCTCCGCTGGATTCGCTCTACAGTTTCGACAAAGCCGTCAT	185	
Db	61	NKIALELCITFLATALGAVLLLPFSIISINVELLSLRPNYYIQWLNSLIHLNVLNLFPP	120	
Qy	186	TGCCAAATCGCGTGTTATCTTCGGTGGCGGATCTACCGGTATCACAGTCTCCGATCAGG	245	
Db	121	NLSLIFLMPAYFTTESEGFAGSRKGVLRVYEVTVMLMLLTLLVLGMVWVASAIVDKN	180	
Qy	246	CCTGATCTTTACCAACACACACTGCGGATCGGTGTCTATCCAGAGCCCAAGCACGCTGA	305	
Db	181	ANRESLYDFWEYLYPYLSCISPLGVLALLVCPPLGLARMFVTGKLLVKPRLLEDBEQ	240	
Qy	306	TCACGACATCTCGCGCATGGTTTCGTTCTCCACGAGTGGGTGAGAGCTCTCCGATCC	365	
Db	241	LYCSAFEEAALTRICNPTSCWLPLOMELLHROVLALQTVRVLLEKRRKASANQNLGYP	300	
Qy	366	GGGTCTTTCCGTGAAGTATCTGCGCAAGATCGTGAAGGTAAACGACAGGTAGAAGACA	425	
Db	301	LAMLCLLVLTGSLVLIVAIHILELLIDEAAPRGMQGTSLGOWSFSLGSGFAGVATQVVL	360	
Qy	426	GCTCAAGGGTATCACTGACGAGATGAGCGCTCTGCCCAAGCTCAGGAGGTATGCCAAGA	485	
Db	361	FYLMVSSWVGFSPLFRSLRPRWHDTAMTQIIGNCVCLLVLSALPVSFRTLGLTRFDL	420	
Qy	486	ACTGGCCAAAAAGAAATGCAGA	509	

Db 421 LGDFGRNWLGNFYVFLYNAFA 444

RESULT 13

US-09-923-876-5908

Sequence 5908, Application US/09923876

Patent No. US20020013958A1

GENERAL INFORMATION:

APPLICANT: Laigudi, Raghunath V.

APPLICANT: Kamigaki, Laura Y. (Ito)

APPLICANT: Sherman, Bradley K.

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING

FILE REFERENCE: PL-0012-1 CON

CURRENT APPLICATION NUMBER: US/09/923, 876

CURRENT FILING DATE: 2001-08-06

PRIOR APPLICATION NUMBER: 09/298,329

PRIOR FILING DATE: 1999-04-21

PRIOR APPLICATION NUMBER: 60/085,331

PRIOR FILING DATE: 1998-05-05

NUMBER OF SEQ ID NOS: 6332

SOFTWARE: PERL Program

SEQ ID NO 5908

LENGTH: 268

TYPE: DNA

ORGANISM: Zea mays

FEATURE:

NAME/KEY: misc\_feature

OTHER INFORMATION: Incyte ID No. US20020013958A1 700457955H1

NAME/KEY: unsure

LOCATION: 13, 50, 69, 179

OTHER INFORMATION: a, t, c, g, or other

US-09-923-876-5908

Query Match 1.6%; Score 34; DB 10; Length 268;

Best Local Similarity 48.4%; Pred. No. 0.98;

Matches 91; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 436 ATCACTGACGAGATGGAGCGTCTCGCAAGACTCAGGAGGTATGCCAAGAACTGCCCAAA 495

|||||

Db 63 ATCACTNCGCAACCATGCGCGGCGAGCAGCGGCGGTGCGTACTGTCCAG 122

QY 496 AAGCAAAATGACAGAGCAACCACTGTCATCGTAGAGCCCTTCTATTCACAAAGAA 555

|||||

Db 123 AAGGACGAGGACATACAGATGATGCTCCCGCCGCGTCCACCTTGGCACCACCAAGAAATGC 182

QY 556 TACTTCTCTCACTGCTACGATGATTCAAGGACGTTCTGATGGTATTTGCTCTCCCGCAGC 615

|||||

Db 183 GATTTCAGACGGAGCGCTATGCTTTAAGCGCGTTCCGACGGCATCTACATCATCATC 242

QY 616 TCTCTAGG 623

|||||

Db 243 TTGGCAAG 250

RESULT 14

US-10-152-724A-22/c

Sequence 22, Application US/10152724A

Patent No. US20030082714A1

GENERAL INFORMATION:

APPLICANT: LITTLE, Melissa

APPLICANT: HOLMES, Gregory

APPLICANT: KOLLE, Gabriel

APPLICANT: YAMADA, Toshiya

APPLICANT: GEORGAS, Kylie

APPLICANT: WILKINSON, Lorine

TITLE OF INVENTION: No. US20030082714A1el Nucleic Acid and Polypeptide

FILE REFERENCE: P22378

CURRENT APPLICATION NUMBER: US/10/152,724A

CURRENT FILING DATE: 2002-05-23

PRIOR APPLICATION NUMBER: Australian App No. US20030082714A1PQ 4348

PRIOR FILING DATE: 1999-11-26

NUMBER OF SEQ ID NOS: 24

SOFTWARE: PatentIn version 3.1

SEQ ID NO 22

LENGTH: 50000

TYPE: DNA

ORGANISM: Homo sapiens

US-10-152-724A-22

Query Match 1.6%; Score 34; DB 9; Length 50000;

Best Local Similarity 56.1%; Pred. No. 23;

Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 362 TTCGCGGTCTTCCGTGAAGTATCTGCGCAAGATCGTGAAGTAAAGGACAAAGGTAGAG 421

|||||

Db 18679 TTCGAAACGTTTGATGAGATCAGAGCCAGACATGAAAGAGAGAGAGAGGAAAAAG 18620

QY 422 GACAGCTCAAGGTTATCACTGACGAGATGGAGCGTCTGCGCAAAAGCTCAGGAGG 475

|||||

Db 18619 GCCAGCAGAACTTTACCGCTGAATGACAAGGATGCTGAGCTCTGTTGAGGTGG 18566

RESULT 15

US-09-808-880-1

Sequence 1, Application US/09808880

Publication No. US20030027287A1

GENERAL INFORMATION:

APPLICANT: Betlach, Mary C.

APPLICANT: Shah, Sanjay Krishnakant

APPLICANT: McDaniel, Robert

APPLICANT: Tang, Li

TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE

FILE REFERENCE: 30062-20029.00

CURRENT APPLICATION NUMBER: US/09/808,880

CURRENT FILING DATE: 2001-03-14

PRIOR APPLICATION NUMBER: US/09/428,517

PRIOR FILING DATE: 1999-10-28

PRIOR APPLICATION NUMBER: 60/120,254

PRIOR FILING DATE: 1999-02-16

PRIOR APPLICATION NUMBER: 60/106,100

PRIOR FILING DATE: 1998-10-29

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 50937

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA

US-09-808-880-1

Query Match 1.6%; Score 33.6; DB 9; Length 50937;

Best Local Similarity 47.2%; Pred. No. 31;

Matches 102; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 1377 CGACAAGCTCCCCGATATATTCAAGATGTAATCGACAAGAAATTCAGAAGCGGACACGAA 1436

|||||

Db 44477 CGCGAGATCTACGACCGCGGTACCGGGCGGCGGCAAGGACTACGCCGCGAGCGGAA 44536

QY 1437 GAAGTATGCACTTCGTATTTCGACAAGAGTGTGTTCTTATAGCGACAAGTTCCATGC 1496

|||||

Db 44537 GGAGTGGCGACCTCGTGGCGGACCGGGTGGCGGACCGCTCTCCCTCTGGAGGTGGC 44596

QY 1497 CATGCTCAAGTCCATGGACAAGAAAGTTTGCCAAAGGCTATCGAGAAAGATCCGCGAGT 1556

|||||

Db 44597 CTGCGGACGGGCGCACCTGCGGCACTTCGCCAGGCTCTTCGACGACGCGCCCGGCTCT 44656

QY 1557 AGAGCTTCCAAAGCGCTAATAGTGTGCTGCTGCGGC 1592

|||||

Db 44657 CGAACTGTCCGCGACATGCTGGACATCGCCCGCTC 44692

Search completed: May 23, 2003, 07:01:24

Job time : 741 secs,



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 23, 2003, 02:45:01 ; Search time 2048 Seconds  
(without alignments)  
16915.110 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaaatgaattaaaaag.....aagagctgaagtgtatctaa 2139

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*\*  
1: em\_estba:\*\*  
2: em\_esthum:\*\*  
3: em\_estin:\*\*  
4: em\_estnu:\*\*  
5: em\_estov:\*\*  
6: em\_estpl:\*\*  
7: em\_estro:\*\*  
8: em\_htc:\*\*  
9: gb\_est1:\*\*  
10: gb\_est2:\*\*  
11: gb\_htc:\*\*  
12: gb\_est3:\*\*  
13: gb\_est4:\*\*  
14: gb\_est5:\*\*  
15: em\_estfun:\*\*  
16: em\_estom:\*\*  
17: gb\_gss:\*\*  
18: em\_gss\_hum:\*\*  
19: em\_gss\_inv:\*\*  
20: em\_gss\_pin:\*\*  
21: em\_gss\_vrt:\*\*  
22: em\_gss\_fun:\*\*  
23: em\_gss\_mam:\*\*  
24: em\_gss\_mus:\*\*  
25: em\_gss\_other:\*\*  
26: em\_gss\_pro:\*\*  
27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	110	5.1	537	17	BH388186
c 2	110	5.1	591	17	BH392840
c 3	108.2	5.1	500	17	BH386991
c 4	103.8	4.9	418	17	BH375201
c 5	72.4	3.4	426	17	BH400391
c 6	66.4	3.1	697	17	BH375688

7	55	2.6	398	10	AW774807	AW774807	EST333958
c 8	49.2	2.3	1005	17	BH164253	BH164253	ENTTG63TF
c 9	42.8	2.0	555	17	BH400866	BH400866	AG-ND-125
10	42.2	2.0	692	9	AI405287	GH25242-5	
11	41.2	1.9	721	17	BH391233	BH391233	AG-ND-141
12	39.4	1.8	822	17	BH371846	BH371846	AG-ND-162
13	39	1.8	606	9	AL728310	AL728310	AL728310
c 14	38.6	1.8	687	17	BH448937	BH448937	BOG2L67TR
15	38.4	1.8	184	10	AW375157	QV2-CT012	
16	38.4	1.8	466	12	BF426788	df72a12-y	
17	38.4	1.8	491	10	AW148212	da13h05-x	
18	38.4	1.8	505	14	BQ163407	BQ163407	952078A08
19	38.4	1.8	540	14	BQ035218	BQ035218	952078A08
20	38.4	1.8	600	14	BQ163454	BQ163454	952078F07
c 21	38.4	1.8	895	12	BF784434	602110820	
22	38	1.8	465	10	AV628406	AV628406	
23	38	1.8	497	10	AV621707	AV621707	AV621707
24	38	1.8	587	13	BI247683	BI247683	602959191
c 25	38	1.8	1091	10	BE282103	BE282103	601100927
26	37.8	1.8	422	10	BB851067	BB851067	BB851067
27	37.8	1.8	455	14	BQ241220	BQ241220	TaeO5007G
c 28	37.8	1.8	461	9	AL372451	AL372451	MLBA51A03
c 29	37.8	1.8	680	13	BI309279	BI309279	EST530689
c 30	37.8	1.8	773	12	BG444863	BG444863	EST506482
31	37.6	1.8	422	10	AV629946	AV629946	AV629946
32	37.6	1.8	927	12	BG437679	BG437679	602489932
33	37.4	1.7	367	13	BG983576	BG983576	TL5-CN006
34	37.2	1.7	318	10	AV644201	AV644201	
35	37	1.7	348	12	BF717547	BF717547	fd46902-y
36	37	1.7	606	14	BQ806505	BQ806505	WHE3579-H
37	37	1.7	838	14	BQ671037	BQ671037	AGENCOURT
38	37	1.7	1016	14	BQ059614	BQ059614	AGENCOURT
39	36.8	1.7	668	9	AU070770	AU070770	AU070770
40	36.6	1.7	313	12	BE804382	BE804382	sr79a10-y
41	36.6	1.7	436	10	AV637236	AV637236	AV637236
42	36.6	1.7	471	10	AV640279	AV640279	AV640279
43	36.6	1.7	478	10	AV643425	AV643425	AV643425
44	36.6	1.7	497	10	BE552689	BE552689	946083H03
45	36.6	1.7	509	12	BF252639	BF252639	EST419901

#### ALIGNMENTS

RESULT 1  
BH388186/c  
LOCUS BH388186 537 bp DNA linear GSS 11-DEC-2001  
DEFINITION AG-ND-124P3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-124P3,  
DNA sequence.  
ACCESSION BH388186  
VERSION BH388186.1 GI:17334327  
KEYWORDS GSS.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.  
REFERENCE 1 (bases 1 to 537)  
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.  
Direct Submission of BAC-end sequences from Anopheles gambiae  
Unpublished (2001)  
Other GSSs: AG-ND-124P3.TR  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: biloftus@tigr.org  
This clone is from an A. gambiae BAC library (ND-TAM) provided by  
F.H. Collins and sequenced by The Institute for Genomic Research  
(TIGR). The BAC library was generated from A. gambiae PEST strain  
DNA. All DNA was extracted from newly hatched first instar larvae

to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 For  
Class: BAC ends.

## FEATURES

source Location/Qualifiers

1. .537  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-124P3"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 153 a 112 c 94 g 178 t  
ORIGIN

Query Match 5.1%; Score 110; DB 17; Length 537;  
Best Local Similarity 55.9%; Pred. No. 1.5e-20;  
Matches 209; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 527 TCGTAGAGCCCTTCTATCCAAACAACTACTTCCTCATCGTCTACGATGTATTCAAGG 586  
DB 374 TGGTGAAGGATTTCTTCAAAGGAATAGTTTACTACTTTGTATTCAGGATTTCAAAG 315  
QY 587 ACGTTCGTATGGTATTGCTCTCCAGCTCTGTAGTAAAGTTCGAGGCGATACGGACA 646  
DB 314 ATGTTCTGTGTAGTACTCCGCGCTCATCTATTGGTAATATGGTGGTACTACTGATA 255  
QY 647 ACTGGATGTGGCCCGCTACACAGGCGGACTTCAGCGATTCGCGGTGTATCCGCGTCCG 706  
DB 254 ACTGGGAATGCCAAGACACACAGGAGACTTCTCTGTTTCCGCTGTATCTGATCAAAA 195  
QY 707 ACACCGCGCGCGGATACAGCAAGACATACACCTATAACCCGCTTACTTCGCTG 766  
DB 194 ATGCAATCTGCTGTAATATTCAGCTAACACGTTCTTTAAAGCCTAAGCATCATTTAC 135  
QY 767 CCGTATCCATCAAGGCTACAAGGCTGACGACTATGCGCATGACCATCGTTTCCCGGGCA 826  
DB 134 CAATTTCACTAAAGGAATAAGCTGGTATTTGCAATGATGTAGGATATCCTGGTA 75  
QY 827 GTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAACAGCAACAATC 886  
DB 74 CAACAAACAGATACCTACTTCTTTCCGAAATCGAACAATGGTAAGCAAAAGATTACCCGG 15  
QY 887 CTCGTATCGAAGTT 900  
DB 14 CATGGTTGAAGCT 1

## RESULT 2

BH392840/c  
LOCUS BH392840 591 bp DNA linear GSS 11-DEC-2001  
DEFINITION AG-ND-162P14.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162P14  
, DNA sequence.

ACCESSION BH392840  
VERSION BH392840.1 GI:17338981  
KEYWORDS GSS.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.

REFERENCE 1 (bases 1 to 591)  
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.  
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae  
JOURNAL Unpublished (2001)  
COMMENT Other GSSs: AG-ND-162P14.TF

Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 For  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1. .591  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-162P14"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 170 a 122 c 105 g 194 t  
ORIGIN

Query Match 5.1%; Score 110; DB 17; Length 591;  
Best Local Similarity 55.9%; Pred. No. 1.6e-20;  
Matches 209; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 527 TCGTAGAGCCCTTCTATCCAAACAACTACTTCCTCATCGTCTACGATGTATTCAAGG 586  
DB 374 TGGTGAAGGATTTCTTCAAAGGAATAGTTTACTACTTTGTATTCAGGATTTCAAAG 315  
QY 587 ACGTTCGTATGGTATTGCTCTCCAGCTCTGTAGTAAAGTTCGAGGCGATACGGACA 646  
DB 314 ATGTTCTGTGTAGTACTCCGCGCTCATCTATTGGTAATATGGTGGTACTACTGATA 255  
QY 647 ACTGGATGTGGCCCGCTACACGGCGGACTTCAGCGTATTCGCGGTGTATCCGCGTCCG 706  
DB 254 ACTGGGAATGCCAAGACACACAGGAGACTTCTCTGTTTCCGTTTATGCTGATAAAA 195  
QY 707 ACACCGCGCGCGGATACAGCAAGACATAAACCTATAAGCCCGTTCCTTACTTCGCTG 766  
DB 194 ATGCAATCTGCTGTAATATTCAGCTAACACGTTCTTTAAAGCCTAAGCATCATTTAC 135  
QY 767 CCGTATCCATCAAGGCTACAAGGCTGACGACTATGCGCATGACCATCGTTTCCCGGGCA 826  
DB 134 CAATTTCACTAAAGGAATAAGCTGGTATTTGCAATGATGTAGGATATCCTGGTA 75  
QY 827 GTACGGATCGCTACCTCACTTCTTGGGGTGTGGAAGATCGTATCGAAACAGCAACAATC 886  
DB 74 CAACAAACAGATACCTACTTCTTTCCGAAATCGAACAATGGTAAGCAAAAGATTACCCGG 15  
QY 887 CTCGTATCGAAGTT 900  
DB 14 CATGGTTGAAGCT 1

## RESULT 3

BH386991/c

LOCUS BH386991 500 bp DNA linear GSS 10-DEC-2001

DEFINITION AG-ND-148L11.TR ND-TAM Anopheles gambiae genomic clone AG-ND-148L11

, DNA sequence.

ACCESSION BH386991 GI:17333133

VERSION BH386991

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

Anopheles.

REFERENCE 1 (bases 1 to 500)

AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

[illegible]

```

VERSION      BH400391.1  GI:17346607
KEYWORDS     GSS.
SOURCE       African malaria mosquito.
ORGANISM     Anopheles gambiae
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
             Anopheles.
REFERENCE    1 (bases 1 to 426)
AUTHORS     Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE       Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL     Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-147H4.TR
             Contact: Brendan J Loftus
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0208
             Fax: 301 838 3543
             Email: b.loftus@tigr.org
             This clone is from an A. gambiae BAC library (ND-TAM) provided by
             F.H. Collins and sequenced by The Institute for Genomic Research
             (TIGR). The BAC library was generated from A. gambiae PEST strain
             DNA. All DNA was extracted from newly hatched first instar larvae
             to minimize the inclusion of DNA from microorganisms that inhabit
             the gut. The DNA is derived from mixed sexes of larvae. The BAC
             library was constructed at Texas A&M University BAC Center
             University, College Station, Texas 77843-2123, USA using a HindIII
             partial digest.
             Seq primer: M13 For
             Class: BAC ends.

FEATURES     Location/Qualifiers
             source
               1..426
               /organism="Anopheles gambiae"
               /strain="PEST"
               /db_xref="taxon:7165"
               /clone="AG-ND-147H4"
               /clone_lib="ND-TAM"
               /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT   111 a 79 c 76 g 160 t
ORIGIN

Query Match      3.4%; Score 72.4; DB 17; Length 426;
Best Local Similarity 57.8%; Pred. No. 1.2e-09;
Matches 149; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

Qy 1756 GGTGCTGGTACAACTATCATACACAGGCAAGGCGTATTGGAGACGAGGATCCTTAAG 1815
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 GGTATCAAGCAAGAAATTACTATACCACAAATGGAAGGTATGATTAAGAAGTACAAAGAAAGT 199
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1816 AGCGATGAGTTTGGCGGTACAGAGAATATCTCGACCTCTTCCGCACCAAAACTATGGT 1875
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 GACCAAGAATTTCGATCTTCCACAGGACTCTTTCATCTTTATAAAGAAAAATTATGGT 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1876 ---CGCTATGCCGAGAGAGGTACGTCATATCGCTTTTCCTATCGAACAGCATCAGC 1932
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 ATGTACAAAGACAAAGAGGGCAACTTCATGTAACCTTCTTCTTAATACGATATTACA 79
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1933 GCGGTAACCTCCGTAGCCCGTATTCGATAAGACGCGCTGATCGGCTTTCGCTTC 1992
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 GGAGGTACTTCAGGTTCTCCAAATTATCGATGGTTACGGAAGACTTATAGGTCCTTGCAATT 19
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1993 GATGGCAACTGGGAAGCT 2010
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18 GACGGAACACAGTGAAGCT 1
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
LOCUS     BH375688/c
DEFINITION AG-ND-120J9.TR ND-TAM Anopheles gambiae genomic clone AG-ND-120J9,
          DNA sequence.
ACCESSION BH375688
VERSION   BH375688.1 GI:17321830

```

```

KEYWORDS     GSS.
SOURCE       African malaria mosquito.
ORGANISM     Anopheles gambiae
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
             Anopheles.
REFERENCE    1 (bases 1 to 697)
AUTHORS     Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE       Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL     Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-120J9.TF
             Contact: Brendan J Loftus
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0208
             Fax: 301 838 3543
             Email: b.loftus@tigr.org
             This clone is from an A. gambiae BAC library (ND-TAM) provided by
             F.H. Collins and sequenced by The Institute for Genomic Research
             (TIGR). The BAC library was generated from A. gambiae PEST strain
             DNA. All DNA was extracted from newly hatched first instar larvae
             to minimize the inclusion of DNA from microorganisms that inhabit
             the gut. The DNA is derived from mixed sexes of larvae. The BAC
             library was constructed at Texas A&M University BAC Center
             University, College Station, Texas 77843-2123, USA using a HindIII
             partial digest.
             Seq primer: M13 Rev
             Class: BAC ends.

FEATURES     Location/Qualifiers
             source
               1..697
               /organism="Anopheles gambiae"
               /strain="PEST"
               /db_xref="taxon:7165"
               /clone="AG-ND-120J9"
               /clone_lib="ND-TAM"
               /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT   214 a 133 c 122 g 228 t
ORIGIN

Query Match      3.1%; Score 66.4; DB 17; Length 697;
Best Local Similarity 50.3%; Pred. No. 9.1e-08;
Matches 197; Conservative 0; Mismatches 186; Indels 9; Gaps 1;

Qy 16 AAAAGTATTCTTCTCGGAGCAGCCCTGCTGTGGTGTCTTCAGGGTAGCCAAAGCCGAC 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 AAAACATCTCTATTAGCAGTCTTACTCCCGCAGTGACGGCTTTTGCTCAACAGCG 416
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 76 AAAGGCATGTGGCTCCCTCAACGAACTCAATCAGGAGAACTCTGGATCGAATCGTGAGCTC 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 GGAGGAATGTGATTCTCTACAGAGTTAAATGAAAAGGA-----AATGAAGGAATTG 365
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 136 GCGTTTACGCTCCGTTGGATTCTCCTACAGTTTCGACAGCGCGTCCATTTGCCAATGCC 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 GGCATGAAGATTCTCGGAAAGATATTTTCAATACTCAAAAACCTAGTATTAAAGATGCT 305
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 196 GTGGTTATCTTCTCGTGGGATGTACCGTATACAGTGTCCGATCAGGCGCTCATCTTT 255
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 GTAGTGAATTCACAGGAGGCTGACTGTGTAATATTTTCTCCAAAAGGTTTGTATTG 245
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 256 ACCAACCCACTCCGGATACGGTGTCTATCCAGACCCAAAGACGCGTGGATCAGCACTAT 315
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 ACAATCACCACCTGTGCTTCGGACAGATTCAAGCTCATTCACATGACAAATGACCTT 185
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 316 CTGCGCGATGGTTTCGTTTCTCGGACGATGGGTGAGAGCTTCGCGATTCGGGTCTTCC 375
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 CTAAGTAATGGTTTCTCGGCAAAAACATGGGAGAAGAACTTCCAAACCCCTGGTGTGTT 125
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 376 GTGAAGTATCTCGCCAAGATCGTCAAGGTAAC 407
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 GTAGATTTCATCCCGATATTAAGAGGTGAC 93
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



Qy	307	CACGACTATCTGCCGATGGTTT	329
Dd	268	CAAAATCATTTAAAAAATGGTTT	290
RESULT 8			
BH164253/c			
LOCUS	BH164253	1005 bp	DNA linear GSS 24-SEP-2001
DEFINITION	ENTT5637F Entamoeba histolytica sheared DNA Entamoeba histolytica genomic, DNA sequence.		
ACCESSION	BH164253		
VERSION	BH164253.1	GI:15737691	
KEYWORDS	GSS.		
SOURCE	Entamoeba histolytica.		
ORGANISM	Eukaryota; Entamoebidae; Entamoeba.		
REFERENCE	1. (bases 1 to 1005)		
AUTHORS	Lofthus,B., Wang,Z., Van Aken,S. and Fraser,C.		
TITLE	Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: bjloftus@tigr.org Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library		
Seq primer:	M13-Forward		
Class:	Shotgun		
High quality sequence start:	15		
High quality sequence stop:	487.		
FEATURES	Location/Qualifiers		
source	1..1005	/organism="Entamoeba histolytica"	
	/db_xref="taxon:5759"		
	/clone_lib="Entamoeba histolytica Sheared DNA"		
	/note="Vector: PHOS1; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."		
BASE COUNT	186 a 314 c 336 g		
ORIGIN			
Query Match	2.3%; Score 49.2;	DB 17;	Length 1005;
Best Local Similarity	56.0%;	Pred. No. 0.012;	
Matches	93; Conservative	0; Mismatches	73; Indels 0; Gaps 0;
Qy	195	CGTGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTGTCCGATCAGGCGCTGATCTT	254
Dd	229	CGCGGTGGTGC GCCCGGTGGTGGTGTACCGGTTCGTTCTGTCGCCGACGCGCTGCTGCT	170
Qy	255	TACCAACCACCATCTCGGATACCGTGGTATTCCAGAGCCAAAACGACGGTGGATCAGGACTA	314
Dd	169	GACCAACCACCATCTGGCGATGGCGTGTATCCAGTACAACAGCTGCSCCGACGACGACCT	110
Qy	315	TCTCGCGGATGGTTCGTTTCTCGACAGATGGGTGAGGAGCTTCG	360
Dd	109	GATCACTGGTGGCTTCATCGCAACGGGCCGTGCCGACGAGGCGCCG	64







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 08:43:01 ; Search time 3929.8 Seconds  
(without alignments)  
5272.840 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MOWKLSILLGALLGASG.....LPMIDKWGCPRLIOELKLI 712

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US10008355/runat\_16052003\_110400\_2535/app\_query.fasta\_1.1230  
-DB=GenEmbl -OPT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=diplom62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355\_@CGL\_1\_1\_2983\_@runat\_16052003\_110400\_2535 -NCPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_jmu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*

29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	907.5	24.4	10689	1	AE004008 Xylella f
C 2	186.5	5.0	253217	2	AC016590 Homo sapi
C 3	175	4.7	10811	1	AE011732 Xanthomon
C 4	147.5	4.0	50372	1	AE014150 Streptoco
C 5	146.5	3.9	10889	1	AE006549 Streptoco
C 6	144.5	3.9	15569	1	AE000733 Aquifex a
C 7	137.5	3.7	162754	8	AL662969 Oryza sat
C 8	133.5	3.6	4004	1	X55799 M.sp 207 al
C 9	133	3.6	10292	1	AE008428 Streptoco
C 10	133	3.6	11362	2	AL449945 Streptoco
C 11	131.5	3.5	14169	1	AE009858 Pyrobacul
C 12	130	3.5	2652	6	A81135 Sequence 1
C 13	130	3.5	2652	6	AR065782 Sequence
C 14	130	3.5	2652	6	E26074 Novel vals.
C 15	130	3.5	9827	1	AE013500 Methanosa
C 16	130	3.5	10029	1	AE010493 Fusobacte
C 17	129	3.5	4141	1	CLOCOLA
C 18	129	3.5	6004	6	BD003707 Polynucle
C 19	129	3.5	11280	1	AE007366 Streptoco
C 20	129	3.5	26703	3	CBRG45011
C 21	128.5	3.5	291150	1	AP003135 Staphyloc
C 22	128.5	3.5	342600	1	AP003363 Staphyloc
C 23	128	3.4	11264	1	AE009964 Streptoco
C 24	127	3.4	193188	2	AC006884 Caenorhab
C 25	126.5	3.4	90348	1	AF497482 Micromono
C 26	126	3.4	3084	1	U62096 Lactobacill
C 27	126	3.4	11925	1	AE013596 Methanosa
C 28	125.5	3.4	52900	1	AE014138 Streptoco
C 29	124.5	3.3	9389	1	AE008276 Agrobacte
C 30	124.5	3.3	11199	1	AE009339 Agrobacte
C 31	124	3.3	36380	3	AF106581 Caenorhab
C 32	123.5	3.3	2142	1	AF039313 Moraxella
C 33	123.5	3.3	10029	1	AE010835 Methanosa
C 34	123.5	3.3	10296	1	U67538 Methanococc
C 35	123	3.3	10181	1	AE001761 Thermotog
C 36	123	3.3	340806	1	AL162752 Neisseria
C 37	122.5	3.3	4716	6	AX415417 Sequence
C 38	122.5	3.3	4731	6	AX413695 Sequence
C 39	122.5	3.3	7009	14	AF296095 Porcine t
C 40	122.5	3.3	7858	1	U35629 Lactococcus
C 41	122.5	3.3	10826	1	SPU49397 Streptococ
C 42	122.5	3.3	239050	1	AL596169 Listeria
C 43	122.5	3.3	299850	1	AP001514 Bacillus
C 44	122.5	3.3	349980	6	AX413017 Sequence
C 45	122.5	3.3	349980	6	AX417045 Sequence

ALIGNMENTS

RESULT 1

```

AE004008/c
LOCUS      AE004008      10689 bp      DNA      linear      BCT 15-JUN-2001
DEFINITION Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.
ACCESSION  AE004008 AE003849
VERSION     AE004008.1 GI:9106961
SOURCE      Xylella fastidiosa 9a5c.
ORGANISM    Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1 (bases 1 to 10689)
AUTHORS    Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrier,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
TITLE       The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
JOURNAL     Nature 406 (6792), 151-157 (2000)
MEDLINE     20365717
PUBMED      10910347
REFERENCE   2 (bases 1 to 10689)
AUTHORS    Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Cartaro,D.M.,
            Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr. A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silva,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovsky-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
            Direct Submission
TITLE       Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES   Location/Qualifiers
            source      1..10689
                        /organism="Xylella fastidiosa 9a5c"
                        /db_xref="taxon:160492"
                        /clone="9a5c"
            gene         complement(119..367)
                        /gene="Xf1871"
            CDS           complement(119..367)
                        /gene="Xf1871"
                        /notes="hypothetical protein; identified by sequence
                        similarity; putative; ORF located using Glimmer/RBSfinder"
                        /codon_start=1
                        /transl_table=11
                        /product="hypothetical protein"

/protein_id="AAF84677.1"
/db_xref="GI:9106962"
/translation="MTMTLKEDKTEIDFQRIQIAKLSETAKIQAEIHEVNAHTOKL
MRETLTLTTESKWYVWVGSSLIAGATAATLFIKLFH"
complement(430..618)
/gene="Xf1872"
CDS           complement(430..618)
/gene="Xf1872"
/notes="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84678.1"
/db_xref="GI:9106963"
/translation="MMEEARATAPQGGLAASDPQRTDAPKKTGLRKRYPTTGSASI
VLURHVLTQQRKASP"
941..1288
/gene="Xf1873"
CDS           941..1288
/gene="Xf1873"
/notes="similar to GI|479340 (percent identity: 36 %/query
alignment coverage: 87.8 %/subject alignment coverage:
90.2 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84679.1"
/db_xref="GI:9106964"
/translation="MSKNKIPNQLTGNWAGFOFINGKLVTPERRITTEWQLRWSLT
CTLARQWKMEEARATAPQGGLAASDPQRTDALQKTVSHKRHAANETASTIQLRQ
VLQROQQRKASP"
1351..1662
/gene="Xf1874"
CDS           1351..1662
/gene="Xf1874"
/notes="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84680.1"
/db_xref="GI:9106965"
/translation="MTMTLKEDKTEIDFQRIQIAKLSETAKIQAEIHEVNTHTOEM
NAHNEMNARTQEMNARTQKLIETQKVSQSWYVWVGSLIAGATAATLFIKLF
H"
complement(1839..2918)
/gene="Xf1875"
CDS           complement(1839..2918)
/gene="Xf1875"
/notes="similar to SP|P03626 (percent identity: 36 %/query
alignment coverage: 88.0 %/subject alignment coverage:
105.0 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84681.1"
/db_xref="GI:9106966"
/translation="MLNLVTGAPNGKTLIYADWLIRQIEIDKSLVKSGAVPRSFYTD
IEGFDEAVRRLTGVVQSPAPEDWRTTPOGSIYVDEAHMFPAGRPGRSDDPVCDL
DTHRHGGYDLMTFTWPKTIHRLRLVGFHVLNRMGLQTAGLYRMSRAQDDPYDI
HOREKABEEVWKFPPKDRYALYASSTLHTVSHKFRIPKVKNSALSVCVTSVGLVFWH
YSPAPHLSEAAASVAGASGASLRAAPASLSRSSRLVSGMRTYAVLETESAPLSCGV
SSDRSCRCNTDGYQIDMSVVECRLLASPLPNFYHAYVYVTSSSSSAAAVSSSPSSPS
complement(2918..3202)
/gene="Xf1876"
CDS           complement(2918..3202)
/gene="Xf1876"
/notes="similar to SP|P03625 (percent identity: 36 %/query
alignment coverage: 85.1 %/subject alignment coverage:

```

```

86.0 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84682.1"
/db_xref="GI:9106967"
/translation="MFSWLAMLRNLTGQVARGLVAGALVTVVPLIPLVTSALNL
IVKMGISADVNLIALMGFGFALSTIGSAMLTRLALQSLHVGIVKATT"
gene
complement(3195..4412)
/gene="XF1877"
CDS
complement(3195..4412)
/gene="XF1877"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84683.1"
/db_xref="GI:9106968"
/translation="MCSRSEFLYFISALGTLVISFPFLFAQNVPPTVEVPTSVRY
SSVSDGVSATSFARSTALVNGVRYTPVDISATLGLSLAKVAVRRGMAYNYVS
MLSLINGACWVIDELTHEVMSGPALKEIPVGTVAWMQRPQGDHFVSYTPQGLIAP
INAYOSSLPQPLVTSQSPSSVSSERWLYHLEGGCEYIGHLSKTDQSPDYSSGLPP
SVILTDLGLQVRSDPSTVNAVLTDSQTGAVALITPEIVSALNKLRSLEDELKASHAP
DQPSGGASPSPPSSQWAMPSCFSCWASVVCDFIDWKSDEFLLKPLPLPDPVPVVDKL
PQAKTWSGEGEGACPSPTTPIEFSGYKTSVEISYQPFCDFAALRPVIVATILA
AYTAGGPRGVKNV"
gene
complement(4481..4702)
/gene="XF1878"
CDS
complement(4481..4702)
/gene="XF1878"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84684.1"
/db_xref="GI:9106969"
/translation="MKSYIDRGIAFFSLLSTSSLVFAADAAGAFDAGAAVATLGAIS
GAVALICAAKLAPAAISVGNKWLKGAIFG"
gene
complement(4724..4924)
/gene="XF1879"
CDS
complement(4724..4924)
/gene="XF1879"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84685.1"
/db_xref="GI:9106970"
/translation="MAQQLMSLYCTQYDVEARTCSQQAWMVPPSLPPISEDVRIIL
PHIVMCFVLVAGFHLFTVVRD"
gene
complement(4931..5236)

```

## Alignment Scores:

```

Pred. No.:      3.58e-53      Length:      10689
Score:          907.50      Matches:      239
Percent Similarity: 49.21%      Conservative: 135
Best Local Similarity: 31.45%      Mismatches: 291
Query Match:      24.40%      Indels:      95
DB:              1          Gaps:      21

```

US-10-008-355-2 (1-712) x AE004008 (1-10689)

```

Qy      1 MetGlnMetLysLysLeuSerIleLeuLeuGlyAlaLeuLeuLeuGlyAlaSerGly 20
|||||  ||| |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10499 ATGCGGTTTAAACCTGTTGCGCTTTCCGTTTGTAGCATATTGATAACTGTGCGACTCC--- 10443

Qy      21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAsp 40
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10442 ---ACTCATGCGGGCGAAGGATGTGGGTTCACAAACAGCTT---CCGGAAATTCGTGCGC 10389

```

```

Qy      41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10388 CCTCTAAGCAGCGGTGGCTACAACCTGTCCACCAACAATTTGTCAAATTTGACAGGTGAC 10329

Qy      61 SerIleAlaAsnAlaValAlaIlePheGlyCysThrGlyIleThrValSerAsp 80
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10328 CCCATGGGGCAGGTGGTCTCTCTC-----GGCAATTCACCCGAGCTGCTGTTTCACCA 10275

Qy      81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10274 GAAGGACTGGTGATCATAACCATCATTCGCTTATGGTGGCATTCAGTTGAACCTCTACG 10215

Qy      101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluLeuLeuPro 120
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10214 CCTAAAAGAACTTAATCAAGAAGGGTTCAATCGTGTGACTCAGGCTGATGAAGTCAGT 10155

Qy      121 Ile---ProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysVal 139
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10154 CGGGTCCGAATGCGGTATTATTACGTGCTTGAGGAGATTACCGATGTCACCGCTCAAGCT 10095

Qy      140 GluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluVal 159
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10094 AAAGCTGCTCTGGCTGCTGCGGTAAAGCCCTTCAAGCGTACGACGGCATTGGAGACA 10035

Qy      160 Cys-----GlnGluLeuAlaLysGlyGluAsnAlaAspGluAsnGlnLeuCysIleVal 177
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10034 TTCAGCAACAAGAAATCGCCAATCGCAGGAA---GAGCAGGGTTATCGCTGCCAATTC 9978

Qy      178 GluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspVal 197
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9977 TTCAGTTTCGCTGGTGGAAATACCTCAGAGATGTTCAAGAACCTGGAGATCAAGGACGTG 9918

Qy      198 ArgMetValPheAlaProPheSerValGlyLysPheGlyGlyAspThrAspAsnTrp 217
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9917 CTTTGGTTTACGCCACCGCAGCGGAGTGTGGTAACTTCGGGGCGGAGCTGCATAACTGG 9858

Qy      218 MetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsn 237
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9857 ATGTGGCGCGCTCATACCGGTGACTTTTCTTCTACCGCTGCCTACGTCGGTAAAGACGCT 9798

Qy      238 ArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaVal 257
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9797 AAACCCGCGAGCTTCTCTAAAGAAAACATACCTACCGCTCCCAACACTGGCTGAAATTT 9738

Qy      258 SerMetGlnGlyTyrLysAlaAspTyrAlaMetThrIleGlyPheProGlySerThr 277
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9737 AGCATCAACCATTAGGTGATGGTGACTTCGTGATGGTTCGCTGGTACCCAGGAGCTACC 9678

Qy      278 AspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGlnAsnAsnProArg 297
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9677 AATCGTTATGCTTTGGTG-----GCCGAATTGAGAATACTGCTCAC 9636

Qy      298 IleGluValArgGlyIleLysGln-----GlyIleTrpLysGluAla 311
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9635 TGGACTTATCCGGTATGGCCAGCATTTCAAGATCTCATTTGTTGATTGAGGCGGCT 9576

Qy      312 MetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsn 331
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9575 AGCAACAGATCCAGACATTCAGTGAATATCGGACACTTTAGCTGGATTGAATAAC 9516

Qy      332 TyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg 351
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9515 ACATCGAAAAACTTTGATGGTCAATTGGATGGTCTCCGACGCAATTAATGCTATAGGTGAG 9456

Qy      352 LysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGly----- 367
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9455 AAACAGAGTGAAGAACACAGCAGTATTACCTTGGTTGAAACACGACGAGGATACGCTGACAC 9396

Qy      368 LysSerAlaValTyrGlyAspValLysSerLeuGluLysAlaTyrLysGluGlyAla 387
|||||  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9395 GAAGCATTCGCGCATCAACATTCGTTGATCTGCAGACAAATATAC----- 9348

```









Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A.,  
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,  
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos  
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
 Kitajima, J.P.  
 Comparison of the genomes of two Xanthomonas pathogens with  
 differing host specificities  
 Nature 417 (6887), 459-463 (2002)  
 22022145  
 12024217  
 2 (bases 1 to 10811)  
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Parah, C.S., Furlan, L.R.,  
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida  
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,  
 Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,  
 Chambergo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,  
 Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,  
 Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,  
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite  
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A.,  
 Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,  
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,  
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.R., H.A.,  
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,  
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos  
 Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
 Kitajima, J.P.  
 Direct Submission  
 Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de  
 São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900,  
 Brazil

#### FEATURES

##### Location/Qualifiers

```

1..10811
/organism="Xanthomonas axonopodis pv. citri str. 306"
/strain="306"
/db_xref="taxon:190486"
/note="pathovar: citri"
complement(293..1315)
/gene="tdh"
/note="XAC1022"
complement(293..1315)
/gene="tdh"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="threonine 3-dehydrogenase"
/protein_id="AAM35905.1"
/db_xref="GI:21107162"
/translation="MKALVKKREANKGIWLEQVVPVTPGPNVLEIKLEKTAICGDLHI
YLWDSQRTIEPTIGIEFVGRVAELGSAVTGYOIGRVSAGHIVGHCNCRGG
RPHLCPNTVGVNVNGAFVAVMPASNLWPIPDQIPSELAAPFDPVGNAAHCALEF
DVIGEDVLITGAGPIGIIAGICKHIGARNVVVDVNDPRKLADMGATRVVNSKT
SLKDVMDLMEGFDFGLMSGNSRAFNMDLMDYHGGKIAMLGIMPRGAGCDWDKII
FKGLTVQGIYGRKKMYETWKMTQLVLSGFPLHKVLTHQLPIDDQKQFDLMBEGRAGK
VVLWSN"
complement(1840..4260)
/gene="feca"
/note="XAC1023"
complement(1840..4260)
/gene="feca"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="tonB-dependent receptor"
/protein_id="AAM35906.1"
/db_xref="GI:21107163"
/translation="MFHEPLTPCVAKRRLCTVRPPSAVPEQDSPMTSARRFHVTPAL
ALATLLPVISOAATPADPAAPAAPEAALTAADAAANTOLDAINVVSQGSTQVORISO
ODIEQLTPGSAFAKVELPGVQOSADPCTYEMSTQVTLHGDSRLGYLDGIPL
GNMSGVYNGLHITRAISENLGVSVEIAQAGALGATSNLTGGMYSADPQITPG
ARLVGTGSDATRRITFVRADTGDRLDSAYLSYANASTDKWKGYGDTQSEANLKTYY

```

```

OMGDNRLSLFVDTSRREYDYMDSLTSQALGWNYYDLOPDWATVQMARAYQNTG
TSCVANGYPOSGLAGLPSYSLWLSAYVAGGLRRDLNLAGLSGTFTVGGATLDAAGYV
HCNRCGEQWVPYVTSQAIPVSMRTTDYGLDRFCGTSAKWSMGNHDLVGCWAEANA
RTTQRNYPALCNEGYTSLYVYEAQTFPRDFLQRYTQTRMLYQDTRVLLDRLT
LNYGAKALSTTTAARSLVPTTALAAGRIARAEDNFPOLGINTKLDERODFLYASKNKI
AAFGTFPATSOAAFSRSTLEPEQSQTVOVYQDOAQLSDADAYFTFKFSNRLLT
TSPCTAVQTCISILSNVGAHAGADLALMRPTEGLSWLNSLWNSRYSRYQDYLNNG
VYATSKDVGVGIPALMFSSASYQIGNRLDLDDGKYVKRKYITFLNDQVSPSYWLFNA
GARXDFGRGVGADVVALNLSNLTDKRYFASTGTNGYVSPDPDGYNOTMVMVGAAPQY
FMSLDVKE"
4369..6450
/gene="XAC1024"
4369..6450
/gene="XAC1024"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="non-hemolytic phospholipase C"
/protein_id="AAM35907.1"
/db_xref="GI:21107164"
/translation="MTSSRRDRFLKRVAAALTAAGALPSSSTGRALALANARTGLRD
LEHVILMQENKSPDHYFGLRGVGRGDPRALQRLDGHAVWSQPTDGRRLVLPFAFD
SQNTCAPLIKSLDHSKAGQDPAWAEYDAWPKYKGLTMGYFQRHDIPIYHALAD
AFTICDGYICSLHGPTNPNMYLFTGTSGPSVNGVQAQAVTNADGNNTADMARCKPG
YAALOWTYYAORLOAAGVDWRYQVDFNFGCNSLAYFSYHRLDRDDEYLRARACVP
GACAGNAATQAEHLVAATAADVRANRLPOYSWITPTAYCEHPAPAYGCSLVARL
IDALTANPEVWAKTALIINDENDGFFDHPVAPALDARMGSRNVDTHGEYDGVPI
GLGIRPMLVISLSPTRGWNVSQVDFDTSVLRLERFGEVAPENISPMRRVSGDLTS
VFDERKPDLSALSALPSYDDYDRAARTVDRKPLPSAPAAATMPROEQORPARALPYA
LQVHARVQGAQVQLQVNSGAAAFNVNSAAGGPPWYTVLPGTQLDDAPICATH
EGAYALRVHGPNGFLREFAGVPRASAPSAAPVEARQEGDALVLEIGNAGORACTVO
LRALDYADPSARTLSLAAGRETIRLAASDHVYDLVVEQPGSFRRLAGHLETVGR
PSRDPAEGRV"
6617..6766
/gene="XAC1025"
6617..6766
/gene="XAC1025"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM35908.1"
/db_xref="GI:21107165"
/translation="MRPNLFAVSVLATLAVASQAQAGSGMWVPOQLPEIAGPPQAGL
RLSPR"
6791..7054
/gene="XAC1026"
6791..7054
/gene="XAC1026"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM35909.1"
/db_xref="GI:21107166"
/translation="MGAVVALGGCTASVSPQGLVVVTHHCAYGAIQLNSTAQKNLIK
DGFNAVRPADELSAGPSARIYVLDITDVTAPAKAAMATPVRR"
7051..7236
/gene="XAC1027"
7051..7236
/gene="XAC1027"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM35910.1"
/db_xref="GI:21107167"
/translation="MMDAQGLVLGLAFQGNWESSVSNWIFDPMTRMTAVDGRYLWRI
MTEVAPAPQLIKELGVR"

```

```

gene      7329. .7973
/genes="pgmA"
/notes="XAC1028"
CDS       7329. .7973
/genes="pgmA"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="phosphoglycerate mutase"
/protein_id="AAM35911.1"
/db_xref="GI:21107168"
/translation="MRLLARHGCTPNWAEGRYQGOIDPLSPVGGQARALGERLHS
LQIDRVASPLSRAQAAGAASQALQIDADLOEIAHGWEGLLASEINDKDP
LRAMREPTVLPQGESLRQVLDRSWRGLRAADGLGAHDTLLVVAHDVNRVILC
KIIGLPLSLRWSFRQAPTTLINLEGGDVEHLEVRVNDCAHHTPFGEAKHRL"
8242. .9540
/genes="folC"
/notes="XAC1029"
CDS       8242. .9540
/genes="folC"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark; dihydrofolate
synthase"
/codon_start=1
/transl_table=11
/product="folylpolyglutamate synthase"
/protein_id="AAM35912.1"
/db_xref="GI:21107169"
/translation="MNATDLSLAWLAIQOHPASAIAMGLERVAVRLQIAAPAKH
VIVVGTNGKSTVATIEAIGRAAGWKVGYATSPHLIRNERNVRIDQASDAOLVDA
FAAVEAARGDTALTTFEFTLAAALWLFQOASALELAVLEITGLGRLDAVNIVDSDAVI
Alignment Scores:
Pred. No.:      0.0237      Length:      10811
Score:          175.00      Matches:      32
Percent Similarity: 75.00%      Conservative: 16
Best Local Similarity: 50.00%      Mismatches:  16
Query Match:     4.71%      Indels:      0
DB:              1          Gaps:          0

US-10-008-355-2 (1-712) x AE011732 (1-10811)
QY 646 GlyAsnSerGlySerProValPheAspGlyAsnGlyArgLeuIleGlyLeuAlaPheAsp 665
Db 7033 GGCAACTCCGGTTCGCGGTGACGACGCGCAGGCAAGCTGTCGCGCTTCGAC 7092
QY 666 GlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 685
Db 7093 GGAANTGGAGTGGTGAGCAGCACTGATCTCGACCGCGCAATGACCCGATGATC 7152
QY 686 SerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysProArgLeu 705
Db 7153 GCCGTGACGCGCGTATCTGCGTTGATCATGACCGAGGTGCTCGCGCCACACGCTG 7212
QY 706 IleGlnGluLeu 709
Db 7213 TTGAAGGAACCTC 7224

RESULT 4
AE014150
LOCUS      AE014150      50372 bp      DNA      linear      BCT 18-JUL-2002
DEFINITION Streptococcus pyogenes MGAS315, section 15 of 37 of the complete
genome.
ACCESSION AE014150 AE014074
VERSION   AE014150.1 GI:21904382
KEYWORDS
SOURCE     Streptococcus pyogenes MGAS315.
ORGANISM   Streptococcus pyogenes MGAS315
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE  1 (bases 1 to 50372)
AUTHORS   Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
Direct Submission
Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
Hamilton, MT 59840, USA
FEATURES
source
1. .34426
Location/Qualifiers
/organism="Streptococcus pyogenes MGAS315"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/focus
/notes="group: A"
34427. .50372
/organism="Streptococcus pyogenes phage 315.1"
/db_xref="taxon:198538"
208. .1206
/genes="SpyM3_0653"
208. .1206
/genes="SpyM3_0653"
/notes="best non-GAS blastp hit: gb|AAK99779.1| (AE008470)
ABC transporter substrate-binding protein - unknown
substrate [Streptococcus pneumoniae R6]"
/codon_start=1
/transl_table=11
/product="putative ABC transporter substrate-binding
protein"
/protein_id="AAM79260.1"
/db_xref="GI:21904383"
/translation="MNKSLIATLLVITVIGLSLKGVSKEKRDLANOPNITIGL
QFVTHEALDDIKRGIEDQLKQKQKQNVVIVKVMNAEGDSKIQTSRQVQSGSDIV
IGIATPAAGLAAGKSIDPVMASVSDPVGSRVLMQDQPEANVTGLSNKVPVKQTID
LMKKLPHVKTGVLIASNEDSLQSVKFRRLAKGQVIVSYAVPSTNEVPATMSV
MLGKVDVAFIPQDNTIASAFSSVMTTSKAAKIPVYTSVDRMVEKGGLAALSONOYDLG
VQTANOVLLIKIKGRVVDVPKVDVIGQPLINKNVAAELGAIINREDFEPSFIEN"
1157. .1375
/genes="SpyM3_0654"
/notes="best blastp hit: gb|AAK33913.1| (AE006546)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM79261.1"
/db_xref="GI:21904384"
/translation="MLLTGKIFLQLLLIRINTKTCIDIKGQEKIKTKNTLTENLDAKSD
NRLEDYTVFIIFSFKMLYFRHVKNLAF"
1551. .2420
/genes="SpyM3_0655"
1551. .2420
/genes="SpyM3_0655"
/notes="best non-GAS blastp hit: gb|AAK99780.1| (AE008470)
ABC transporter membrane-spanning permease - unknown
substrate [Streptococcus pneumoniae R6]"
/codon_start=1
/transl_table=11
/product="putative ABC transporter membrane-spanning
permease"
/protein_id="AAM79262.1"
/db_xref="GI:21904385"
/translation="MIVSSVSQGLIWLGLGIYLTFRILNFFDWTTEGSPFLGGAVA

```

VTALSGWNFFLSTLLGMSLALAGFLTGILLYTKGKMTPTLAGILVMTSNCISIMLWVM  
GRANLGHDKRQDCPLFPISDILNSLITGLITVIVISVLVYFTYMLGQATAGDN  
KDMAKSGIINTDMWENVGLVNSLIASLALGALVSQODGYADVSKGIGVIVIGIASIIV  
GEVLYSGLTLRLERLAINVIGSIYQLFISVIMLGENTSYLKLISALVALCLMIPV  
VKERFFKGVRLTR

gene

2417. .3175

/gene="Spym3\_0656"

CDS

2417. .3175

/gene="Spym3\_0656"

/note="best blastp hit: gb|AAK33915.1| (AE006547) putative  
ABC transporter (ATP-binding protein) [Streptococcus  
pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="putative ABC transporter (ATP-binding protein)"

/protein\_id="AM79263.1"

/db\_xref="GI:21904386"

/translation="MTKIIELINATVDVNGFEDAKTILDNVLTIVYEHDFLTILGN  
GAKSTLFNIAGTSLTRQGIIRILGQDVTVPAEKRALYLSRVFQDPKMGTPRMTV  
AENLLIARQGERKSLASRKITEHLASFEDLVKRTNGLEKHLLETAGLLSGGORAL  
SLLMATLKKPALLLDDEHTAALDPKTSLSMQLTDERVTDKGLTALMITHHMEDALTY  
GNRLVYMKDGNIIKOLNQMEKEQITIDYQLFD"

gene

3379. .5040

/gene="Spym3\_0657"

CDS

3379. .5040

/gene="Spym3\_0657"

/note="best non-GAS blastp hit: gb|AAK74765.1| (AE007370)  
metallo-beta-lactamase superfamily protein [Streptococcus  
pneumoniae TIGR4], andgb|AAK9342.1| (AE008432) Conserved  
hypothetical protein [Streptococcus pneumoniae R6]"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AM79264.1"

/db\_xref="GI:21904387"

/translation="MTDIKMIALGGVREYKKNFYLVBEINDSMFILDAGLKYPENEQIG  
VDLVPNLVDVIEKNGVQIGFISLGHADAIGALPYLLAEVSAPVFGSELTIELAKLF  
VKNSSTFKFNNFVDSDEIEFKDGLVSFFRTTHSIPESMGIVIGTDKGNIVYTG  
FKFDQAAREGYQTDLLRAEIGKEGVLALLSDSNVATSNQDIASESVEGEMDSVTD  
ADGRVIAVASNLVRIQOVFDSATAGHREVLTGTDAENIVRTALRLEKMLTDERL  
LIRPKDKFEDHELIIILEAGRMGEPINSLQKMAAGHRVYQIQEGDLVYIVTPSTA  
KEAWARKVENLITKAGSVKLTIONLRVSGHANGRDQLLMLNLLKPYLFPVQGEYRD  
LAARAKLAEVGEFFIPENIHILKRGDITMVLNDEGFLHEGGVPASDVMDIGNAIDVNI  
VLDRKVLSEDDGIFVIAITVSKKEKRIISKAKVNTRGVYVVKKSHDILRESAELVNT  
VGNVKKDTFDMGELGNVRDLSKFLFEQTKRRPALLPVMEVR"

gene

5173. .5958

/gene="esta"

CDS

5173. .5958

/gene="esta"

/note="best blastp hit: gb|AAK33917.1| (AE006547) putative  
tributylin esterase [Streptococcus pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="putative tributyrin esterase"

/protein\_id="AM79265.1"

/db\_xref="GI:21904388"

/translation="MASIAIEYHSVLGMERKVNVIYPDQSEIPKKDQGDPIVLVYL  
LHGKGNENSMOKRTATIERLLRHTNLIVMPSTDGLGWYTPATYGLNRYALSOELPOV  
LAAPFPNWKORKETFFVAGLSMGYGAFAFKWALKSNRFSYASFGALDFSPETLLGN  
LGELAYWQGVFGDFDDLDKHLKNNVAESDGRKTFYAWCYGVEDFLFATNKAIAADF  
QAQGLDIDYHKHGKHEWYNNQOLEVLEMLPYNQKEERLS"

gene

5988. .6314

/gene="Spym3\_0659"

CDS

5988. .6314

/gene="Spym3\_0659"

/note="best blastp hit: gb|AAK33918.1| (AE006547)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AM79266.1"

/db\_xref="GI:21904389"

gene

6360. .8267

/gene="Spym3\_0660"

CDS

6360. .8267

/gene="Spym3\_0660"

/note="best blastp hit: gb|AAK33919.1| (AE006547) putative  
ABC transporter (ATP-binding protein) [Streptococcus  
pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="putative ABC transporter (ATP-binding protein)"

/protein\_id="AM79267.1"

/db\_xref="GI:21904390"

/translation="MILOQNKLERSFSGDVLFQNLISLOVDERDRLALVCPNAGKST  
LLKLLGVEETPSGEVNTKDLTSLVLAQNSREESDKTIIEEMLKVFALRQDEKRUR  
OMDMATVSGQDLTRIMMDYDLAERFQOGFTYEAETKATLNGFKFDESMWQMTI  
AELSGGONTRLAKMLLEKPELVLDDEPTNHLIDETIAEKENYLANOQCALITVSHD  
RYFLDKVATVTLDTLPHGLDRYVYNGVSRFMTLKAELVABEQFDQKQKIAKLEDFV  
OKNIVRASTTKRAQARRKOLEKTERLDKPTSAKSAHMTFOADKPCSNVVLTVKKAAT  
GYNQHVLSFEPINLDVHKLDAIIVGPNIGKSTLTKSVIGQIPIFKGEVYKGVANVEIG  
IYDOTSHLUTSSNVILEELWODEFTTPEVDIRNLGAFLESGDDVKKSVAMLSGGKRA  
KLLAKLSMNNFLVLDPTNHLIDSDSEVENALIDFDGTLFLVSHDRYFNRLAT  
KVLLETNGSTLYLDGYDYLEKKAELAELELARAAGETVETKEASATDYLOKANOK  
ERRLRTRYBEIEARLETIERIGATQEDMHASNDTAQLIAWQKEDQLDQEQEALME  
EWETIAFOIES"

gene

8552. .9520

/gene="acoa"

CDS

8552. .9520

/gene="acoa"

/note="best blastp hit: gb|AAK33920.1| (AE006547) putative  
acetoin dehydrogenase (TPP-dependent) alpha chain  
[Streptococcus pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="putative acetoin dehydrogenase (TPP-dependent)  
alpha chain"

/protein\_id="AM79268.1"

/db\_xref="GI:21904391"

/translation="MVTVSKEQHLDMFLKMERIREFDSRINKLVRRGFVQGMTHFSVG  
Alignment Scores:

Pred. No.:	12.4	Length:	50372
Score:	147.50	Matches:	171
Percent Similarity:	32.47%	Conservative:	105
Best Local Similarity:	20.12%	Mismatches:	232
Query Match:	3.97%	Indels:	343
DB:	1	Gaps:	44

US-10-008-355-2 (1-712) x AE014150 (1-50372)

Qy 35 AsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeu 54

Db 30183 AATGAAGAAAACGAGATATCTTAGAGGATATTTGTTTAACTTACCTTATTTGAA--- 30239

Qy 55 TyrSerPheAspLysProSerIleAlaAsnAlaValIlePheGlyGlyCysThr 74

Db 30240 -----GATAGGGAGATGATTAGGAAAGACTTAAACATATGCTCAC----- 30281

Qy 75 GlyIleThrValSerAspGlnGlyLeu-----IlePheThrAsnHisCysGlyTyr 92

Db 30282 -----CTCTTTTGATGATAGCGTGTGATGAAACAGCTTAAACGTCGCCCGTTACTGTTGG 30335

Qy 93 GlyAlaIle-----GlnSerGlnSerThrVal 101

Db 30336 GGACGTTTGTCTCGAAAATTGATTAAATGGTATTAGGGATAAGCAATCTGGCAAAACA 30395

Qy 102 AspHisAspTyrLeuArg---AspGlyPheValSerArg----- 113

Db 30396 ---TTAGATTTTGTGAAATCAGATGGTGTTCCTCCCAATCGCAATTTTATGCAGCTGATCCT 30452

QY 114 -----ThrMetGlyGluGluLeuProIleProGlyLeuSer----- 125  
 DB 30453 GATGATAGTTTGACATTAAAGAACACATTCAAAAGACACAAAGTGTCTGGCAAGGCGAT 30512  
 QY 126 -----ValLysTyrLeuArg 130  
 DB 30513 AGTTTACATGAACATATTTGCAAAATTTAGCTGGTAGCCCTGCTATTAATAAAAGGATTTT 30572  
 QY 131 LysIleValLysValThrAsp-----LysValGluGlyGln----- 142  
 DB 30573 CAGACTGTAAAGTTGTTGATGATTTGTCACAACTAATGGGGCGGCATAAACCCAGAAAT 30632  
 QY 143 -----LeuLysGlyIleThrAspGluMet 150  
 DB 30633 ATCGTTTATGAAATGGCACGTGAAATATCAGACAACTCAAAAGGCCAGAAAATTCGCGT 30692  
 QY 151 GluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLys-----LysGlu 167  
 DB 30693 GAGCGTATGAACCGTATTGAAGAAGGAATAAAAGAACTAGGAAGTGATATTCTTAAGGAG 30752  
 QY 168 AsnAlaAspGluAsn-----GlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyr 186  
 DB 30753 TATCCTGTGAAACACACTCAATTA-----CAAAATGAAAG 30788  
 QY 187 PheLeuIleValTyr-----AspValPheLysAspValArgMetValPhe 201  
 DB 30789 CTCATCTCTATTATCTCCAAATGGAAGACACATGTATGTGGACCAAGAATTAGATATT 30848  
 QY 202 AlaProSerSerValGlyLysPheGlycylcysThrAspAsnTrpMetTrpProArg 221  
 DB 30848 ----- 30848  
 QY 222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlu 241  
 DB 30849 -----AATCGTTTAAGTGAT 30863  
 QY 242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGly 261  
 DB 30864 TATGATGTCGATCAC-----ATTGTTCCACAAAGT 30893  
 QY 262 TyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu 281  
 DB 30894 TTCCTTAAAGACGAT-----TCAATAGACAAATAAGTCT 30926  
 QY 282 ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArg 301  
 DB 30927 TTAACGGCTTCTGATAAAATCGTGTAAATCGGATAACGTTCCAAAGTGAAGAGTAGTCT 30986  
 QY 302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLys 321  
 DB 30987 AAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAAG 31046  
 QY 322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341  
 DB 31047 TTT-----GATAATTTAAACAAAGCTGAACGTGGA----- 31076  
 QY 342 GlyLeuAlaArgLeuAspValIleGly-----ArgLysArgAlaGluGluArgAla 358  
 DB 31077 GGTGTTGAGTGAACTTGATAAAGTTGGTTTTATCAACGCCCAATGTTGTAACACTCGCCA 31136  
 QY 359 PheAlaAspTrpIle-----ArgLysAsnGlyLys----- 368  
 DB 31137 ATCACTAAGCATGTGGCAAAATTTTGGATAGTCGCATGAATACTAATAATACGATGAAAT 31196  
 QY 368 ----- 368  
 DB 31197 GATAAACTTATTCAGAGGTTAGAGTGATTAACCTTAAATCTAAATAGTTTCTGACTTC 31256  
 QY 368 ----- 368  
 DB 31257 CGAAAAGATTTTCCAATTCATATAAAGTAGCTGAGATTAAACATTTACCATCATGCCCATGAT 31316  
 QY 368 -----ArgTyrAlaGluAsnGlyGlnLeuHisIleAl 636  
 DB 368 -----ArgTyrAlaGluAsnGlyGlnLeuHisIleAl 636

DB 31317 GCGTATCTTAATGCGCGTGTGGAACTGCTTTGATTAAAGAAATATCCAAAACTTGAATCG 31376  
 QY 369 SerAlaValTyrGly-----AspValLeuSerSerLeuGluLysAlaTyr 383  
 DB 31377 GAGTTTGTCTATGTGTATTAAAGTTTATGATGTTCTGTAATAAGTATGCTAAGTCTGAG 31436  
 QY 384 LysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu---PheGly 402  
 DB 31437 CAGCAATAGCAAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAATCTCTTC 31496  
 QY 403 GlyThrGluVal-----ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAs 421  
 DB 31497 AAAACAGAAATTTACACTTGCAAATGGAGAGATTCGCAACGCCCTCTA-ATCGAAACTAA 31555  
 QY 421 PheAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLe 441  
 DB 31556 TGGGGAACACTGGA-----GAAATGTCTGGGATAAAGGGCGAGATTTT-----GCCAC 31603  
 QY 441 uAspArgLysValLeuPro---AlaMetLeuAspIleValArgArgArg----- 456  
 DB 31604 AGTCGCAAAAGTATTGTCCATGCCCAAGTCAATATTGTCGAAGAAACAGAAAGTACAGAC 31663  
 QY 457 -----IlePro-----AlaAspLysLeuProAspIle 465  
 DB 31664 AGCGGATTTCTCAAGGAGTCAATTTTACCAAAAAAGAAATTCGGACAAAGCTT----- 31715  
 QY 465 ePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheVa 485  
 DB 31716-----ATGTCTGTAATAAAAGACTGGGATCCAAAAAATATGTGGT----- 31757  
 QY 485 lPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMetAs 505  
 DB 31758 -TTTGATAGTCCAACGGTAGCTTATTCAGTCTCTAGTGTCTTCTAGGTGGAAAA---GG 31813  
 QY 505 pLysGluLysPheAlaLysAlaIle----- 513  
 DB 31814 GAAATCGAAAGAGTATAAATCCGTTAAAGAGTTACTAGGATCAACAATAATGGAAGAAG 31873  
 QY 514 -----GluLysAspProAlaValGluSerLysSerValIleAlaAlaAlaArgAl 531  
 DB 31874 CTCCTTTGAAAAAGATCGGATGTAGCTTTTGAAGACTAAA-----GGATATAGGA 31924  
 QY 531 aileGlnAlaAspAlaMetAla-----AsnAlaTyrAlaIleGluLysG1 546  
 DB 31925 AGTTAGAAAAGACTTAATCAATTAACCTACCTAAATATAGTCTTTTGGAGTTAGAAAAAG 31984  
 QY 546 yLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArg-----AlaLeuPr 564  
 DB 31985 TCGTAAACGGATGTGGCTAGTGGCGGAGAAATGCAAAAAGGAAATGAGTAGCTCTGCC 32044  
 QY 564 oSerAsp---AlaAsnPheThrMetArgMetSer---TyrGlySerIleLysGlyTyrG1 582  
 DB 32045 AAGCAATATGTGAATTTTTTATATTAGTAGTACATTGAAAGTTGAAAGGTT---AG 32101  
 QY 582 uProGlnAspGly-----AlaTrpTyrAsnTyrHisThrGlyLysGlyVa 598  
 DB 32102 TCCAGAAGATAACGAACAAAAACAATTTCTTGTGGAGCAGCAT-----AACCATTA 32152  
 QY 598 lLeuGluLysGlnAspProLysSerAspGluPheAlaValGln----- 612  
 DB 32153 TTTAGATGAGATTATTGAGCAACATCACTGAATTTTCTAAGCGTGTATTATTTAGCAGATGC 32212  
 QY 612 ----- 612  
 DB 32213 CAATTTAGATTAAGTCTCTTGTGCTGATATAACAACATAGACAGAACCAATACGTGAACA 32272  
 QY 613 ----GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly----- 625  
 DB 32273 AGCAGAAAATATTATTCATTTATTTACGTTCCAGCAATCTTGAGCTCCCGCTGCTTTTAA 32332  
 QY 626 -----ArgTyrAlaGluAsnGlyGlnLeuHisIleAl 636  
 DB 626 -----ArgTyrAlaGluAsnGlyGlnLeuHisIleAl 636



```

/translation="MDKKYSIGLDICTNSVGNVITDEYKYPSPKPKFKVILGNTDRHSIK
KNLIGALDDSETAEKRLKTRARRYTRRNRIICYLQEIFSNEMAKYDLDYFFHRLH
ESFLPDEKKEHHPITFEGNIVDEYAYHERPTIYHLRKLKLDVSTKADARLDLYLALAH
MIKFRGHFLIEGLDNPNDVDKLFITQLVOTYNOLPEENINASGVDAKAILPSARLSK
SRRLNLIAQLPECKNGKLGFLNLIALSLGLTFNFKSNFDLAEDAKIQLSKDYDLDL
NLLAQIGDOYADLFALAANKLSDAILLSDILRVNTEITKAPLSASMKIKRDEHQDLTL
LKAALROQPEKTEKIFFDOSKNGYAGYIDGASQEEFYKFKIPILKMDGFEELLVK
LNREDLRLQRTFDNGSIPIQIHLGELHAILLRQEDFYFLKDNREKIEKILTFRIPY
YVGPLARGNSRAFWMTKSEETIPWNFEEVVDKGAQAQSFIERMTNFKNLPNEKVL
PKHSLLYEFTYVNELVKYVYTEGMRKPAFLSGEOKKAIIVLLFKTNKRVTKOLKE
DYFKKTECDSEIISVEDRENASLGTYHDLKLIKDKDFLONEENEDILEDVLTLT
LPEDREMISERLKYIAHFDKDKVMKOLKRRYTGWRSLKRLINGIRDSQSKTILDP
LKSDFGNARNFOLIHDDSLTFKEDIQAKVQSGDLSHEHANLAGSPAIKAGIILQF
KVVDLVMKNGRLYLVLQNGRDMYVDQELDNRLSDYDHDVFPQFKDQSDINVK
HPVNTQLQNEKLYLYLQNGRDMYVDQELDNRLSDYDHDVFPQFKDQSDINVK
LTRSDKNKSNVPESEVVKMKYWROLLNAKLITQKFDNLTKAERGGISELDKA
GFIKROLVETROIKHAVALDLSRMNTKYDENDKLIREVKITLKSILYSDRKFQF
YKVELNNYHHAADAVLNAVGTALIKKYPKLESEFVYGDYKVDYVRKMIASEQBIG
KATAKFTFYKSNMFNFKETITLANGELIRKRLPIETNGETGEIYWDKGRDFAIVRVLS
MPQVNTIVKTEVOTGSKESILPKRNSDKLIARKDMDPKKYGGFDSPTVAYSVLV
AKVEGKSKLASAGELQNEALPSKYVNFLYLASHYEKLKGVKSPEDNOKOLFVBOH
ELENGRKMLASAGELQNEALPSKYVNFLYLASHYEKLKGVKSPEDNOKOLFVBOH
KHYLDRIEQISEFSKRVILADANLDKVLASVYNKHRDKPIREQAENIHLFTLNLGA
PRAFKYFDITIDRKRTYSKEVLDATLIHQISITGLYETRIDLSQLGDD"
8677..9346
/gene="SpY1047"
CDS
/gene="SpY1047"
/notes="Best Blastp hit = pir||E81299 hypothetical protein
Cj1522c [imported] - Campylobacter jejuni (strain NCTC
11168) >gi|6968946|emb|CAB73942.1| (AL139078) hypothetical
protein Cj1522c [Campylobacter jejuni]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33937.1"
/db_xref="GI:13622194"
/translation="MAGWRTVVVNTSHSKLSYKNNHLIFKDAYKTELHLSIEDILLLE
TTDILVLMRLVRLVDVIFDCKRLPTAMLPFYGRHDSSLOLGKQMSERTVK
SOWWTTIAQILNQSCYLGAGSYFEKQSIMLYHGLEFDPNSRNGHAARTYFNTL
FGNDFSRDLHPINAGLDYGTLLSMFAREVVSQMTQFGLKHAQNFQNFASDI
MEPPRLVDKIYVENRNPFPKIKRELFLTLFSDTSYNGKEMYLTLNLSIDYKVKVKA
LNNEGKVPFEFRI"
9343..9884
/gene="SpY1048"
CDS
/gene="SpY1048"
/notes="Best Blastp hit = emb|CAB83918.1| (AL162753)
hypothetical protein NMA0629 [Neisseria meningitidis]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33938.1"
/db_xref="GI:13622195"
/translation="MSYRMRMTLMDPMTDFAERKAYKPKFKFLLSGFIMHQFSI
YSKLLNNTANNAMIGRLRHNPNKGNITLLTVTEKQFARMYILHGERNNCIANSER
LVPLGFAFDES"
9874..10536
/gene="SpY1049"
CDS
/gene="SpY1049"
/notes="Best Blastp hit = pir||T03910

```

Alignment Scores:

Pred. No.:	2,26	Length:	10889
Score:	146.50	Matches:	172
Percent Similarity:	31.98%	Conservative:	103
Best Local Similarity:	20.00%	Mismatches:	223
Query Match:	3.94%	Indels:	363
DB:	1	Gaps:	44

US-10-008-355-2 (1-712) x AE006549 (1-10889)

```

Qy 35 AsnGluGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeu 54
Db- 6395 AATGAAGAAATGAAGATATCTAGAGGATATGTTTAAACATGACCTTATTGAA--- 6451
Qy 55 TyrSerPheAspLysProSerIleAlaAsnAlaValIlePheGlyGlyCysThr 74
Db- 6452 -----GATAGGGAGATGATTGAGGAAGACTTAAACATATGCTCAC----- 6493
Qy 75 GlyIleThrValSerAspGlnGlyLeu-----IlePheThrAsnHisCysGlyTyr 92
Db- 6494 -----CTCTTTGATGATAGAGTGCATGAACACGCTTAAACGTCGCCCTTATAGTGTGG 6547
Qy 93 GlyAlaIle-----GlnSerGlnSerThrVal 101
Db- 6548 GGAGCTTTGCTCGAAATTTGATTAAATGATAGGATAAGCAATCTGCAAAACAATA 6607
Qy 102 AspHisAspTyrLeuArg---AspGlyPheValSerArg----- 113
Db- 6608 ---TTAGATTTTGGAAATCAGATGGTTTGGCAATCGCAATTTTATGACGTCATCCAT 6664
Qy 114 -----ThrMetGlyGluGluLeuProIleProGlyLeuSer----- 125
Db- 6665 GATCATAGTTTGACATTTAAGAAGACATTCAAAACACACAGTGTCTGCACAGGCGAT 6724
Qy 126 -----ValLysTyrLeuArg 130
Db- 6725 AGTTTACATGAACATATTCGCAATTTAGCTGTTAGCCCTGCTATTAATAAAGGTATTTTA 6784
Qy 131 LysIleValLysValThrAsp-----LysValGluGlyGln----- 142
Db- 6785 CAGACTGTAAAGTGTGTGATGAATTTGTCGCAAGTAATGGGGCGGCATAAGCCAGAAAAAT 6844
Qy 143 -----LeuLysGlyIleThrAspGluMet 150
Db- 6845 ATCGTTATTGAATGGCAGCTGAAATTCAGACAACTCAAAAGGCCAGAAAAATTCGCGA 6904
Qy 151 GluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLys-----LysGlu 167
Db- 6905 GAGCGTATGAACGAATCGAAGAGGTATCAAGAATTTAGGAAGTCAGATCTTTAAAGAG 6964
Qy 168 AsnAlaAspGluAsn---GlnLeuCysIleValGluProPheTyrSerAsnAsnGlyTyr 186
Db- 6965 CATCCTCTTGAATAACTCAATTG-----CAAAATGAAAG -7000
Qy 187 PheLeuIleValTyr-----AspValPheLysAspValArgMetValPhe 201
Db- 7001 CTCTATCTCTATATCTCCAAATGGAAGAGACATGATGTGGACCAAGAAATAGATATT 7060
Qy 202 AlaProProSerSerValGlyLysPheGlyAspThrAspThrMetTrpProArg 221
Db- 7060 ----- 7060
Qy 222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlu 241
Db- 7061 -----RATCGTTTAAAGTAT 7075
Qy 242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaValSerMetMetGly 261
Db- 7076 TATGATGTCGATCAC-----ATTGTTCCCAAGT 7105
Qy 262 TyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu 281
Db- 7106 TTCCTTAAAGACGAT-----TCAATAGACAATAAGGTC 7138
Qy 282 ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArg 301
Db- 7139 TTAACGCGTCTTGATAAAATCGTGGTAAATCGGATACGTTCCCAAGTGAAGAAGATGTC 7198
Qy 302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLys 321
Db- 7199 AAAAGATGAAAACCTATTGGAGACAACTTCTTAACCCCAAGTTAATCACCACGTAAG 7258
Qy 322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341

```



```
Db 7259 TTT:::|||||GATAATTTAAAGAAAGCTGAACGTGGA-|||7288
Qy 342 GlyLeuAlaArgLeuAspValIleGly-----ArgLysArgAlaGluGluArgAla 358
|||:::|||||
Db 7289 GGTGTGAGTGAACGTGTAAGAGTGGTTTATCAACGCCCAATGGTTGGAACCTGCACAA 7348
|||:::|||||
Qy 359 PheAlaAspTrpIle-----ArgLysAsnGlyLys-----368
:::|||||
Db 7349 ATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAT 7408
|||:::|||||
Qy 368-----368
Db 7409 GATAAACTTATTCGAGAGGTTAAAGTGAATACCTTAAATCTAAATAGTTTCTGACTTC 7468
|||:::|||||
Qy 368-----368
Db 7469 CGAAAGATTTCCAAATTCATATAAGATAGCTGAGATTAAACAATTACCATCATGCCCATGAT 7528
|||:::|||||
Qy 368-----368
Db 7529 CGGTATCTAAATGCGTGGTGGAACTGCTTTTGATTAAAGAAATATCCAAAACCTTGAATCG 7588
|||:::|||||
Qy 369 SerAlaValTyrGly-----AspValLeuSerSerLeuGluLysAlaTyr 383
|||:::|||||
Db 7589 GAGTTTGTCTATGTGTGAFTATAAGTTTATGATGTCTGTAATAATCATGCTTAAGTCTGAG 7648
|||:::|||||
Qy 384 LysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu---PheGly 402
|||:::|||||
Db 7649 CAGAAATAGCAAGCAACCGCAAAATATTTCTTTACTCTATATCATGCAACTTCTTC 7708
|||:::|||||
Qy 403 GlyThrGluVal----ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAs 421
|||:::|||||
Db 7709 AAACAGAAATATACACTGTGCAATGGAGAGATTGCAAAACCCCTCTA-ATCGAAACTAA 7767
|||:::|||||
Qy 421 PALAHISLADLYLLELysSerLeuAspLysTyrLysAspTyrLysAspTyrLeuProSerLe 441
|||:::|||||
Db 7768 TGGGAAACTGGA-----GAAATGTGTGGATAAAGGGCGAGATTTT-----GCCAC 7815
|||:::|||||
Qy 441 uASpArgLysValLeuPro----AlaMetLeuAspIleValArgArgArg-----456
|||:::|||||
Db 7816 AGTGGCAAAAGTATGTGCCATGCCCAAGTCAATATGTCGAAGAAACAGAAAGTACAGAC 7875
|||:::|||||
Qy 457-----IlePro-----AlaAspLysLeuProAspIle 465
|||:::|||||
Db 7876 AGCGGATTCCTCAAGGAGTCAATTTTACCAAAAGAAATTCGGCAAGCTT-----7927
|||:::|||||
Qy 465 ePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheVa 485
|||:::|||||
Db 7928-----ATTGCTCGTAAAGAAAGACTGGGATCCCAAAAAAATATGTGTGT-----7969
|||:::|||||
Qy 485 lPheAspLysSerValValProTyrSerAspLysPheHisAla-----499
|||:::|||||
Db 7970 -TTTGATAGTCCACGGTAGCTTATTCAGTCTAGTGTCTAGGTGGGAAAAGGAA 8028
|||:::|||||
Qy 500-----MetLeuLysSer-----MetAspLysGluLys 508
|||:::|||||
Db 8029 ATCGAAGAAGTTAAATCCGTTAAAGAGTTACTAGGATCACATTAATGGAAGAAGTTC 8088
|||:::|||||
Qy 508 sPheAla-----LysAlaIleGluLysAs 516
|||:::|||||
Db 8089 CTTTCAAAAAATCCGATTGACTTTTATAGAGCTAAAGGATATAGGAAGTTAAAAAAGA 8148
|||:::|||||
Qy 516 pProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAl 536
|||:::|||||
Db 8149 CTTAATCAATTAACCTACTCTAAA-----8170
|||:::|||||
Qy 536 aMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGly 556
|||:::|||||
Db 8171 ----TATAGTCTTTTGTAGTTAGAAAACGGTCGTAACGGATGTGCTAGTGCCTGCGGAGA 8226
|||:::|||||
Qy 556 uMetTyrProGlyArg-----AlaLeuProSerAsp----AlaAsnPheThrMetArgMe 573
|||:::|||||
```

---

```
Db 8227 ATTACAAAAGAAATGAGCTGGCTGCCAAGCAATATATGTGAATTTTATATATTTAGC 8286
Qy 573 tSer---TyrGlySerIleLysGlyTyrGluProGlnAspGly-----AlaTrp 588
|||:::|||||
Db 8287 TAGTCATTTAGAAAGTTGAAGGT---AGTCCAGAGATTAACGACAAACAATTCCTT 8343
|||:::|||||
Qy 588 pTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGly 608
|||:::|||||
Db 8344 TGTGGAGCAGCAT-----AAGCATTTATAGATGAGATTATTGAGCAAAATCAGTGA 8394
|||:::|||||
Qy 608 uPheAlaValGln-----612
|||:::|||||
Db 8395 ATTTTCTAAGCGTGTATTTTAGCAGATGCCAATTTAGATAAAGTCTCTAGTGCATATAA 8454
|||:::|||||
Qy 613-----GluAsnIleLeuAspLeuPheArgTh 621
|||:::|||||
Db 8455 CAACATAGACACAAACCAATACGTGAACAGCAGAAATATATTATTATTATTACGTT 8514
|||:::|||||
Qy 621 rLysAsnTyrGly-----Ar 626
|||:::|||||
Db 8515 GACGAATCTTGGAGCTCCGCTGCTTTTAAATATTTTATATATTTGATACAACAAATTGATCGTAAACG 8574
|||:::|||||
Qy 626 gTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly 645
|||:::|||||
Db 8575 ATATACGTCTACAAAGAAGTTTATAGATGCCACTCTTATCATCATCAATCCATCCTGTT 8632
|||:::|||||
RESULT 6
AE000733 AE000733 15569 bp DNA linear BCT 25-MAR-1998
LOCUS Aquifex aeolicus section 65 of 109 of the complete genome.
DEFINITION Aquifex aeolicus section 65 of 109 of the complete genome.
ACCESSION AE000733 AE000657
VERSION AE000733.1 GI:2983720
KEYWORDS
SOURCE Aquifex aeolicus.
ORGANISM Aquifex aeolicus.
REFERENCE
1 (bases 1 to 15569)
Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Auja,M.,
Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus
JOURNAL Nature 392 (6674), 353-358 (1998)
MEDLINE 98196666
PUBMED 9537320
REFERENCE
2 (bases 1 to 15569)
Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Auja,M.,
Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
Direct Submission
Submitted (25-JUL-1997) Diversa Corporation, Genomics, San Diego,
CA 92121
TITLE Putative indicates no similarity to known proteins
Hypothesis indicates similarity to a protein of unknown function.
FEATURES
Location/Qualifiers
source
1. .15569
/organism="Aquifex aeolicus"
/strain="VF5"
/db_xref="taxon:63363"
/complement(199..1497)
/gene="purA"
/notes="aq_1290"
/complement(199..1497)
/gene="purA"
/codon_start=1
/transl_table=11
/product="adenylosuccinate synthetase"
/protein_id="AAC07286.1"
/db_xref="GI:2983724"
/transl_table="MEKLIITLGAQWDEGKGIKIVLLSEHFIDITVYQGGSNAGHTVV
VNSQKFTLHLLPTGILHEHVKGVIAGMWVDLVLHKEVKNLEKGIYKRYERLIFSDR
AHLVPMYKLLDLSLFEKGIKIVLLSEHFIDITVYQGGSNAGHTVVVNSQKFTLHLLD
NLDVFNKICEKVFCEKFDIDINQIYEEQLRYEERFENVDLLRFFNTKQSVLFEQA.
```



```

gene
CDS
/translation="MKIIITGEPVGVKTTLLVKKIVERLGRKRAIGFMTVEVRDPETKRK
TGRRIITTEGKKIIFSKFTSKLGSYGVNVQVFEELAIPILEIRAYREAKKDRRKV
IIIDEIGKMELEFSKKFDLVRQIMHDPNVNVVATIPIRDVHPLVKLEIRLPGAVLIEL
TPENRDVILEDIILSLER"
2131..4734
/gene="alaS"
/note="aq_1293"
2131..4734
/gene="alaS"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC07294.1"
/db_xref="GI:2983732"
/translation="MKIIITGEPVGVKTTLLVKKIVERLGRKRAIGFMTVEVRDPETKRK
TGRRIITTEGKKIIFSKFTSKLGSYGVNVQVFEELAIPILEIRAYREAKKDRRKV
IIIDEIGKMELEFSKKFDLVRQIMHDPNVNVVATIPIRDVHPLVKLEIRLPGAVLIEL
TPENRDVILEDIILSLER"
2131..4734
/gene="alaS"
/note="aq_1293"
2131..4734
/gene="alaS"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC07293.1"
/db_xref="GI:2983731"
/translation="MKKEETRTGIGLKDVSLRLVSESPLLPVGVGFVCEYGPVRYLK
IDNEVYKLEISVEKLNKLSVCNIEQVQVKKVVEELRYALPEGVFTIAKVANGIKLR
ELSLTEETSVGKGIIDEDREDLQEMPAYVKNRETILGTIREVILLELMYSS"
5306..7738
/gene="clpC"
/note="aq_1296"
5306..7738
/gene="clpC"
/codon_start=1
/transl_table=11
/product="ATP-dependent Clp protease"
/protein_id="AAC07290.1"
/db_xref="GI:2983728"
/translation="MFEKFTERARQVILRAREBALELGHSLVLSHILLALIKDEDIP
VLVLSKLGVSPEKIRKAIMQGITKSGSVEVLFPADKARVIEFAVEARILHHQFVGP
EHLILGVREKTLGGLRILSGFLDEYSVRREVLIQLLGEPLPQESQRYVSTNDRFS
EDLTQMARGLKDPVIGVREKIERVIOILVRRKNNPNVLLGDPGVGKTAIVEGLAORI
ANKVEPPLNKRVALDLAALVAGTRYRGQFERLKNILKELEKANNVILFIDELHT
LVGASREGSDASNMULKPALARGEIOVIGATTVDYRKYTEKDGALERRQPVYVEQ
PSPEDTIQIILGLKPFKEEFHVEYTPPEAIEKAVELSVKYITDRNLDPDKAIDVMDKAG
ALVKLKLATQPLPELKEIEEKIELEKEKEKAVKBQDYEKAAKIRDELRLRAKFPETMK

```

```

gene
CDS

```

```

AKWEEKNRPKVTDEDVAEVVARTGVPVKRVKHSDAEKLLHLHIEBELKKRVVQODE
IKAVAKAIRRSRVGLGKRGPIGVFMFLGPTGVGKTKETAKALAEYLFETEDALIFRD
MSYMEKHTYVRLIGAPGVVGEQGLTEAVRRRPYSVLLDEIDEIKAPDVFENIRFL
QIPDDGLTDLGLTRTVDFSNTIIMTSLNIGARLITGTMGEKKQGLLNEEMRKNVL
DOVKRTPSPFELNDELIIVFRPLEKDKDIEQILELQKEINEKLKDNVVRVKLHKSFV
NWIIKEYKPEFGARSUKRALQRHVEDLLAEELIKGLTKGVVEEVEIRVRNDKPYLKPV
7776..10112
/gene="omp"
/note="aq_1300"
7776..10112
/gene="omp"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AAC07292.1"
/db_xref="GI:2983730"
/translation="MKKEFTLALFLGVANAEEIKIIGKNKYPIDELIRGLLSVRK
GSEFSLERIRKDKLYRTGLTKVEAYEEDQNKNGPEVILVVVDLPVIYKIEFEGN
EEFSDDLKLEYLGIETELGEVEYETIEGTSSPAIERLAVMKQLKGLRILTFQEI
SLRRRIEYVKRHGFVGTQVYKLIIPKKGASKLVFIHEGKKIYTSIEIEGKHFSD
RLKGLMELKEPNIFLRLHPARSEETLKEDVKKIEEFYKNEGFLVEVSYVEKKEGP
ARKVIVKESGPYKKEIKLEGTNLEAYSELVDNILLKNERKGRYFRREVIEKIKNR
IREKABIGFLNTSVEERNVNPKEKVSLLKIEGKPVYVKAIRKINGNITREDYVI
RREMVRQENELALKKGIERSKTRIMNLGYIEDVQIEPVRDAMDLLLVKTRFTGO
FSVGLSYNEVTLAGFTIELRKGNFRGTGAGISVSYGSLYRNNAISYTRKWFLLKPV
DLDSADFRIEYDTYVTRTGTFSVALSKSEYRASIGTSIQRVKYSOIDPEASTY
VKQAGRRDSRKIFFTLTRDTRDYLLPTKGLSFVRNSVCGVLLGDEKFFYKFEVEG
AKYPSDYFDTGIIILSKGVGVGEGYGGKKVPIDERFFVGGDSIRGXYGYAGPVG
NITLDPJGAKKKVITSLEMLPVFKMLYFAGFDYGLGADKSDFKLSNFRGGYIGV
VRIITPAPIRIDWAFKTKTPVGTNRSRIHFVLGTFF"
complement(10138..10350)
/gene="cspC"
/note="aq_1303a"
complement(10138..10350)
/gene="cspC"
/codon_start=1
/transl_table=11
/product="cold shock protein"
/protein_id="AAC07291.1"
/db_xref="GI:2983729"
/translation="MSRFGTVKFSKQKGYGFTITREDTNADVVFHFTDIOQEGFKTLQ
KGQKFEVDVVEDTKGPRAKNVRVLGE"
complement(10454..11170)
/gene="hlsA"
/note="aq_1303"
complement(10454..11170)
/gene="hlsA"
/codon_start=1
/transl_table=11
/product="phosphoribosylformimino-5-aminimidazole
carboxamide ribotide isomerase"
/protein_id="AAC07284.1"
/db_xref="GI:2983722"
/translation="MNLKEFTIPADILMEGKAVRLYKGFDFNKVKVYSERPWELAKNFS
DLGKRLHVVLDLEAGKPKNLEIRKIRENFECEVEGGGIRSYEVKALFDGEDID
PVIWITLAYKNKEFLKILENPNRVILAIKQKQKVAIGWKKBETAVSPPEFAKYE
NYPWGLYVTYIERDGSGLGVDEPYEIKKHVKKPVIASGGVSSLEDIKKLYGIVES
VWVGKAYEGRITLEDLQNF"
complement(11167..111922)
/gene="pyrDB"
/note="aq_1305"
complement(11167..111922)
/gene="pyrDB"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase electron transfer
subunit"
/protein_id="AAC07287.1"
/db_xref="GI:2983725"
/translation="MLAEINAEVLENTYISNLYRLILRIPDKILKQIEPGHFAMIKP
SDTYDPMGRARAFVADIEHKAVFYDYVYKGTYLLSKRKPGRKRLTFLPGLRFSY
EGDKHLLIGGVGLAGLTLLAKALRDMGKKVFIAYGGRSRHGLGMVLEWLEEFPEYL

```



Db	5338	AGTTATCTTCGAGCGGAGG-----GAAGAAGCCCTCGAGCTGGGACACACACTACTT	5388
Qy	544	uLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuPr	564
Db	5389	GGGAAGTGAACATATTTTGCTGTCTGATAAAGGATGAGGA-CATACCGCTCCTCGTCG	5447
Qy	564	oSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu---Pr	583
Db	5448	TCAGCAAGCTGGGTGT-----CTCTCCTCGAGAAGTAAGAAGGCCATAATAGGCGC	5498
Qy	583	oGlnAsp-----GlyAlaTTPtyrAsnTyrHisThrGlyLysGly--	597
Db	5499	AAATACTAAGGGAAGCCACTCGGAGAGG-TACTCTTCGCCACCCGACGCAAAAGGTA	5557
Qy	597	-----	597
Db	5558	ATAGAGTTTCAGTTGAAGCGAGAATTCTTCCACCAGATTCTGATAGTCCCGAACAC	5617
Qy	598	-----ValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnG	613
Db	5618	CTACTTATCGAGTGGTTAGAGAAAGACGGGACTCGGTGGAAGATTT-----TA	5668
Qy	613	uAsnIleLeuAspLeuPheArgThr-----LysAsnTyrGlyArgTyrAlaGluAs	630
Db	5669	AGGACTTTGGACTGGGAGTACTCCGTAGGAGAGAGGTACTTCACTACTCGGAGA	5728
Qy	630	nGlyGlnLeuHisIleAlaPheLeuSerAsnAspIleThrGlyGlyAsnSerGlySe	650
Db	5729	CTTCCCTCAAGAAAGCC-----AGAGTACGTA-TCCAC	5763
Qy	650	rProValPheAspLysAsnGlyArg---LeuIleGlyLeuAlaPheAspGlyAsnTrpG	669
Db	5764	TCCTTAAGCTTGACAGGTTCTCAAGGAGCTCACCAATGGCAAGGGAAGAAAGCTCGA	5823
Qy	669	uAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspI	689
Db	5824	CCCAGTAATAGGCAGG-----GAAAGGAATAGAGAGATTATCCAGATACGCT	5874
Qy	689	eArg 690	
Db	5875	AGA 5878	
RESULT 7	OSJN00169/c		
LOCUS	OSJN00169	162754 bp	linear
DEFINITION	Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBa0043L24,		
ACCESSION	AL662969		
VERSION	AL662969.2	GI:21912514	
KEYWORDS	HTG.		
SOURCE	Oryza sativa.		
ORGANISM	Oryza sativa		
REFERENCE	1		
AUTHORS	Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q., Hu, X., Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Weng, Q. J., Zhang, L., Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Zhang, Y. J., Lu, Y., Li, C., Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K., Zhou, B., Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P., Fu, G., Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F., Tu, Y. F., Jia, J., Yin, H. F., Zhang, Y., Cai, Z., Chen, J., Kang, H., Chen, X. Y., Shao, C. Y., Sun, X., Hu, Q. P., Zhang, X. L., Zhang, W., Wang, L. J., Ding, J. C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L., Zhu, F. H. and Hong, G. F.		
TITLE	Direct Submission		
JOURNAL	Submitted (27-JUN-2002) Han Bin, National Center for Gene Research, Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233, CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests: bhan@ncgr.ac.cn		
REMARK	Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC clone: OSJNBa0043L24.		



QY 351 ArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAla 370

Db	2049	GAAGCT	-----TGGGGCGCTAAAGTCGCAAACTACTTTA	2081
Qy	371	ValTyr	-----GlyAspValLeuSerLeuGluLysAlaTyrLysGluGlyAla	387
Db	2082	GGCTATCTTGAACACGGGCACGATTCATTCGCTAGACTTTGGTTTAAAGAAACGGCT	2141	
Qy	388	Lys	-----AlaAsnArgGluMetThrTyrLeuSerGlu	398
Db	2142	CGGTCAATTTGTTTAACGGCAGTTTAGAAGCGGCTAATGCGTCTCTTACTTCGCACGAACACGGC	2201	
Qy	399	ThrLeuPheGlyGlyThrGluValValArgPhe	-----AlaGlnPhe	412
Db	2202	AAATTCGATAATACGCCACACCTCGGGCAGTTTATAGGACCCATGACGAAGAGGGTTC	2261	
Qy	413	AlaAsnAlaLeuAlaThrAsn	-----	419
Db	2262	CTCCATTCTCTCGGGATAAGGAAAGTTTACAAGTCGACGCCACCTCTTCAAGCAACA	2321	
Qy	420	-----ProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLys	435	
Db	2322	GCAAAAGGACAACCTGTGATTACTACGGA-----GAAGACGTTTG3CAAACAGGAGCT	2375	
Qy	436	AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArg	455	
Db	2376	ATAATATATCCACAATACGATAATCGT-----TATGATTTTCGTGGGAT	2420	
Qy	456	ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys	475	
Db	2421	CAAGTGAAGGCAATGAAATTTTGGCGCATTACACCAAAATTTTAAACTTTAGA-----CAA	2477	
Qy	476	GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp	495	
Db	2478	GGTATTTCGAAGCTGTTCGAAA-----GGTGAACGCACGCTGTGTGGAGGATCTGAT	2533	
Qy	496	LysPheHisAlaMetLeuLysSerMetAspLysGlyLysPheAlaLysAlaIleGluLys	515	
Db	2532	AAAGATCAATTCCTCTATTCTCTGCTGACTATCAAGATCATAAAGTGTACGTTGGGTGA	2591	
Qy	516	AspProAlaValGluLeuSerLysSerVal-----IleAlaAlaAlaArgAlaIle	532	
Db	2592	AAGCTAGCAAGAAA---AGTAAAGCGGTGACTTTAACCGCTGGATTTGCACACGCTGTT	2648	
Qy	533	GlnAlaAspAlaMetAlaAsnAla---TyrAlaIleGluLysGlyLysArgLeuPhePhe	551	
Db	2649	GTGACCATGCGCTATTTCGGGTACGGAATATACAGCAACACGAGGAAA-----	2696	
Qy	552	AlaGlyLeuArgGluMetTyrProGlyIArgAlaLeuProSerAspAlaAsnPheThrMet	571	
Db	2697	-----GTAAATTTAACTCTCCAGGAAAAGCAGATCCCGGAACGGTTTTATTGACG---	2744	
Qy	572	ArgMetSerTyrGlySerIleLysGly-----	580	
Db	2748	---GTAGAAGGAGGCAACATAACCGGTGTGCGAAACAGACCGGTGAGTGTGTTGAA	2804	
Qy	581	-----TyrGluProGlnAspGlyAlaTyrTyrAsn	590	
Db	2805	CTAGTACCGGAAAACAATATCCGCATTCTACTATAACGGCGAGACAACTGTGTATAAAAC	2866	
Qy	591	TyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAla	610	
Db	2865	TAC-----GGCGCATGGCTATGCAATGATGTTGGCTCTCTCTTCGCAAACTGGCCA	2915	
Qy	611	ValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGlu---	629	
Db	2916	GTAGGGCAACAGATGTTCCAG-----AAAACCGATGATGCGGTGCGCTATATCATCGTGG	2965	
Qy	630	-----AsnGlyGlnLeuHisIleAlaPheLeuSerAsnAspIleThrGlyGly	646	
Db	2970	CCATTACTGACGGAGGAAAATATCCGCTTCTTGTC---ATGGATATTACCGCGGGT	3026	
Qy	647	AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGly	666	

```

Db 3027 GACCGGCGCAAGACGGCGGAGATAAAGCGTTTACGATTTCTTCCACACAAAGCAATCAA 3086
Qy 667 AsnTrrpGluAlaMetSerGlyAsp-----IleGluPheGluPro---AspLeuGlnArg 683
Db 3087 ATTTGGATCAGCAAGGTTCCGACAAAGTGTACAGGTACGAACCGGTAGATTGCCACGA 3146
Qy 684 ThrIleSerValAspIleArgTyrValLeuPheMetIleAsp----- 697
Db 3147 AAT---ACAGTCCGCAATTCATTACACGGGTGAGATGACTTTGGGCATA 3203
Qy 698 ---LysTrrpGly 700
Db 3204 TGGAAATGGGCG 3215

RESULT 9
AE008428
LOCUS
DEFINITION Streptococcus pneumoniae R6 section 44 of 184 of the complete
genome.
ACCESSION AE008428 AE007317
VERSION AE008428.1 GI:15458055
KEYWORDS
SOURCE Streptococcus pneumoniae R6.
ORGANISM Streptococcus pneumoniae R6
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 10292)
Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
DeHoff,B.S., Estren,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., LaGace,R.,
LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McAhren,S., McHenry,M., McLeaster,K., Mundy,C., Niclas,T.I.,
Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rostek,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
Genome of the bacterium Streptococcus pneumoniae strain R6
J. Bacteriol. 183 (19), 5709-5717 (2001)
21429245
PUBMED 11544234
REFERENCE
2 (bases 1 to 10292)
Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszcak,L., Burgett,S.,
DeHoff,B.S., Estren,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., LaGace,R.,
LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
McAhren,S., McHenry,M., McLeaster,K., Mundy,C., Niclas,T.I.,
Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
Zook,C., Baltz,R.H., Jaskunas,S.R., Rostek,P.R. Jr., Skatrud,P.L.
and Glass,J.I.
Direct Submission
Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES
source
1. 10292
/organism="Streptococcus pneumoniae R6"
/strain="R6"
/db_xref="taxon:171101"
repeat_region 147..341
/rpt_family="BOX element"
/rpt_type="dispersed"
/rpt_unit="Box_ABBBC
353..598
/gene="spr0485"
353..598
/gene="spr0485"
/codon_start=1
/transl_table=11
/label="spr0485"
/product="Hypothetical protein"
/protein_id="AAK99289.1"
/db_xref="GI:15458056"
/translation="MEVMDNIIDSVIPVAEVVDKHPVEVLILVELGFKFLANPLMRN

TVGRKVS LKQGS LAGT PMPDKIVRTLEANGYEVIGLD"
598..1932
/gene="spr0486"
598..1932
/gene="spr0486"
/codon_start=1
/transl_table=11
/label="spr0486"
/product="Conserved hypothetical protein"
/protein_id="AAK99290.1"
/db_xref="GI:15458057"
/translation="MTDERIHILRDLLELHNGASPESVQDRDATFTGVSAIEISLM
EHELMNSDSGVTTFEDYMLCDVHANLFKNAIKGVEVLDTBHPGHPVRFVKEENLALRA
ALIRRLDITYELMEDEMLAEWRKGLVQMGVQFDIHYQKKELEFPFIMERYGH
DSPKVMWGVDDOIRLFOITALAKSLPEVSISSVKEAFEAFATPESMIPEESTIL
LMLLSFTODDNIQIAESDAYGIALIRESEKWPVERQSFIEEKTAEEPVQDITAGG
QVOQVIDPGEHFTITFTPREKAVLDHRHQQAQNGLYLSEQANLLNLHLPMEITFV
NKEDIFQYNDNTPADMEIFKFTPSQVRNVELCHPPKYLDKYKAIMKGLREGTKDKY
EMWFKSRSRGKFPVHITYAAVHDEDEGFGVLEYVDITQPIREYDITDYFRGLE"
1942..2196
/gene="spr0487"
1942..2196
/gene="spr0487"
/codon_start=1
/transl_table=11
/label="spr0487"
/product="Hypothetical protein"
/protein_id="AAK99291.1"
/db_xref="GI:15458058"
/translation="MSYEOEFMKFEFAWVNTQIMDMAHKESOKVYEEDODERAKDA
MIRYESRLDAYQFLGKFKFENFKAGKGFHDLPEGLFGERNY"
2292..2687
/gene="spr0488"
2292..2687
/gene="spr0488"
/codon_start=1
/transl_table=11
/label="spr0488"
/product="Hypothetical protein"
/protein_id="AAK99292.1"
/db_xref="GI:15458059"
/translation="MSKKLNKKQLNRLRSGAFSTVTIKVVDVETKKVVKRAQPSAS
AAGKAVSKYKQEAQVATEKQAKVANSVEDFAANLGLPLDRAKTFYDEGIGKASDFK
NWTEKELLDLKIGGPATIKKLENGIKFK"
3099..3548
/gene="spr0489"
3099..3548
/gene="spr0489"
/codon_start=1
/transl_table=11
/label="spr0489"
/product="Hypothetical protein"
/protein_id="AAK99293.1"
/db_xref="GI:15458060"
/translation="MIGVARENAAEQIKQYQKFTVNISDTSNLAQEQAGFISHQEK
LERLGVHYEISERTQIPLDACPLVDCRVDRIVEEDGICHIFAKILERLVAPELLDE
KGHFKNQLFAPTYFMGDGYQRYRYLDKRVDMKGSFKKARKKDKGN"
3535..4092
/gene="spr0490"
3535..4092
/gene="spr0490"
/label="spr0490"
/translation="MARTELPDKIETELRLVLRVTVADAEIDFYASLPEVAVPAGFP
PVKTEDEIYYLEHILPERNQKNLFGYIGYVKGDKIVGSDVDFNRHEDDVLIEIGY
TLHPDYWGRGYVPEAARALIDLAFKDLGLHKIETLCFGYNLQSKRVAERKUGFTLEARI
RDRKDVQGNRCDSLIYGLLKSEWEE"
4094..5056

```

gene

```

/ gene="spr0491"
4094..5056
/ gene="spr0491"
/codon_start=1
/transl_table=11
/label="spr0491"
/product="Hypothetical protein"
/protein_id="AAK99295.1"
/db_xref="GI:15458062"
/translation="MSDVAEEVSSLSSEKQLRQIDVEAYELNDSIIERLAYLEINNNE
KRIVDIEPTIGETMSVQIPEIOKFNOKIKNMFELFSDVSDFLSIKNKLESKELE
IEADVNRMEIHLSSGKLFVDENENQIKOKYSKSEEPDCIHFASYOYDINFTYRQ
CISLRNYSOHTDLPINEVAVSPDDETVIIDFYIDLDYLLNSFNKKWKLKGBELIKLNQ
ETSKDAIALVKEIFYNALTLYGNYNKLFLKLNHNLTVDIKSKLESKLKHSRYISK
ISKYDLYNPGNYTMSPLAAFAAIEEIIYIELSKIGLVKIVNKS"
5068..7719
/ gene="vals"
/ note="spr0492"
5068..7719
/ gene="vals"
/BC_number="6.1.1.9"
/codon_start=1
/transl_table=11
/label="spr0492"
/product="valyl-tRNA synthetase"
/protein_id="AAK99296.1"
/db_xref="GI:15458063"
/translation="MSKELSSKYNPAEVEAGRYQKWLDADVFKPSGDQKAKPYSIVIP
PNVTKGLHGHAWDTTLODIIIRKRMOGFDTLWLPQMDHAGIATQAKVEERLRGEG
ISRYDIGRESFLTKWEMKDEYATTIKEQKMGKLSVDYSREFTLDEGLSKAVRVF
VDLYKKGVIYRGFEIINWPPARTALSDIEVHKDVEGAFYHNNYLEDGSRALVAT
TRPETMGDVAIVNPEDPRYKDLIGNKILPIANKLIPIVGDHADPEFGTGVVKIT
PAHPNDFLVGSRNTPQVNMINDGTNMLVDFEFSCMDRFEARKAVAKLEIEIGALV
KIEKRVHSVGHRTGVVPEPRLSTQWFMQDLAKNAIQNTDEKDFEYPPRFNDT
FLOMENVMDHWITSLRWGHCQIPAWYNADGMYVGEAPEGWTQDDEVDLPTWFS
ALWPESTMGWPEVDSDEKRYEPTSLVTGYDIIFFWVRMIQSLFTGROPFQNVL
IHGLIIDEQGRKMSKLSGINDPMDVIEKYGADALRWFLNSGAPQDVRFSYEKMDA
SNFNIKNISRIIINNGGLLDVAHNDNKTAVTEGAGNVTDRIWLLNNETIARV
TENFDFEPFVAGHILLYNIIWEEFANMYVELTKEVLYSDNEDDKVITRSLVLLTLDKI
LRLLHSIMPEFIEIFQGAEGSVITVAAYTPNPAFEDLAAHTGVESLKLDIRAVNA
RAENVVAPSKPITILVKTSDSOLEAFENSNVNYIKRTNPHELEIASTIPAPLAMS
VITGAEFLPLDLLAVEELARLEKELAKWQELMDVGKKLSNERFVANAKPEVVOK
EKDQADYQAKYDVTVARIDEMKKLVK"
7774..8544
/ gene="spr0493"
7774..8544
/ gene="spr0493"
/codon_start=1
/transl_table=11
/label="spr0493"
/product="Conserved hypothetical protein"
/protein_id="AAK99297.1"
/db_xref="GI:15458064"
/translation="MMEKVMKHPHYELLNLIGYLAQFDKLFKEFCQCSKSEFYRY
VWSLGTAEYTVGVVKNRDLDFDYEDNNRKGWQKAEVYRERKDLIDMGNEVDHVA
EIVKMLLASEGKTGITTEKPIVTRKFRQLQETGMEANFYFLHEDKKEFGQGGT
DARLYGDGVDQVDQVEYSYLAKEVGIKRSKGRVUTAKEFEKVKFEQSDFLILSVTN
LDIPKLVLNLPKHPFKKNIKEIIIEYRSVEDLY"
8827..9630
/ gene="spr0494"
8827..9630
/ gene="spr0494"
/codon_start=1
/transl_table=11
/label="spr0494"
/product="Hypothetical protein"
/protein_id="AAK99298.1"
/db_xref="GI:15458065"
/translation="MQPTYNIDNPNLISYEAKRDLWRIGFLQKVDNLVPSAYMESLAE
KQSRGELTYEQVYEDATAYHHTIDASTEADLVSLRIVELLSSRRGFSFSPATLLAIHK

```

Alignment Scores:  
Pred. No.: 18.3 Length: 10292

```

Score: 133.00 Matches: 106
Percent Similarity: 33.20% Conservative: 64
Best Local Similarity: 20.70% Mismatches: 164
Query Match: 3.58% Indels: 178
DB: 1 Gaps: 25

US-10-008-355-2 (1-712) x AE008428 (1-10292)

Qy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
|||||
Db 5143 GATGTTTCAAGCCTTCAGCGGATCAAAAGGTAAGCCTTATTCAATCGTTATTTCCACCA 5202

Qy 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
|||||
Db 5203 CCAACCTTACAGGTAAACTT-----CACCTT 5229

Qy 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
|||||
Db 5230 GGT-----CACGCT 5238

Qy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
|||||
Db 5239 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCACAAACCGCATCAAGGTTTGTAT 5295

Qy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
|||||
Db 5296 -----ACCTTTGGCTTCTCGGATGGACACGAGGAGGATGGCCACT 5337

Qy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302
|||||
Db 5338 CAGCTAAGGTTGAGGAGCGCTTCGCTGGTGAGGCAATTTCCCGCTATGACCTTGGTCGT 5397

Qy 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
|||||
Db 5398 GAGTCTTTCTGACGAAAGCTGGGAATGGAAGAC-----GAATATGCCACT 5445

Qy 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
|||||
Db 5446 ACCATAAGGAACAATGGGCAAGATGGGCTCTCTCTAGACATATTCGTGACGCGTTTC 5505

Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgGluGluArg 357
|||||
Db 5506 ACTCTTGACGAAAGTTGTCA-----AAGCTGTTCGTGAAG 5541

Qy 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
|||||
Db 5542 GTCTTTGTGGACCTTTACAGAAAGGCTGGATCTACCGTGGTGAGTTATCATCAACTGG 5601

Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
|||||
Db 5602 GACCCACAGCTCGCACAGCCCTTTCTCATATTGAGGTGATTCACAAGGATGTGGAAGGT 5661

Qy 385 -----GluGlyAlaLysAlaAsnArgGluMetThr 394
|||||
Db 5662 GCCTTCTACCACATGAATTACATCTCGAAGATGTTTCACGCGCCCTTGAAGTTGCTACA 5721

Qy 395 TyrLeuSerGlnThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
|||||
Db 5722 ACTGCTCTCGAGACTATGTTTGGGACGTT----- 5751

Qy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
|||||
Db 5752 GCGATTCCGGTTAATCCAGAA-----GACCCCGCTGAC 5784

Qy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
|||||
Db 5785 AAGACACTTGATTGTTAAACAGCTCATCTTCCA---ATCGCTAATAAATCATCCTCCCAATC 5841

Qy 449 MetLeuAsp-----IleValArgArg 455
|||||
Db 5842 GTTGGAGATGAGCAGCAGACGATCTGAGTTTGGTACTGTTGCGTGAATAATCACACCTGCC 5901

Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475

```







Db	4424	CTGGGACCCAGCAGCTCCACAGCCCTTCTGATATTGAGGTGATTCACAGGATGTAGA	4365
Qy	395	-----GlucylalalysAlaAsnArglume	393
Db	4364	AGGTGCTTCTACCACAGTAATATCATGCTGGAAGATGTTTCACGCGCCCTGAAGTTGC	4305
Qy	393	tThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValargPheAlaGlnPheAl	413
Db	4304	TACAACTGCTCCTGAGACTATGTTTGGGGACGTT	4271
Qy	413	aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLy	433
Db	4270	---GCGGTGCGATTATCCAGAA-----GACCCCGC	4242
Qy	433	sTyrLysasp-----TyrLeuProSerLeuAspArgLysValLeuPr	447
Db	4241	CTACAAAGCACTTGATTGGTAAAAACGTCATCCTTCCA---ATCGCTATAAATCAATCCC	4185
Qy	447	oAlaMetLeuAsp-----IleValar	454
Db	4184	AATCGTTGGAGATGAGCAGCAGATCCTGAGTTTGGTACTGGTGTCTGTAATAATCACACC	4125
Qy	454	gArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysph	474
Db	4124	TGCCACGRTCCAATGACTTCTGGTTGGCCACGTCATACCTGTCACACAGTCACAGT	4065
Qy	474	eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe	494
Db	4064	CATGAACGACGACGACGACATATGATGATGCTTTCGCTTGAATTTTCAGGCATG----	4013
Qy	494	rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGl	514
Db	4012	-GATCGTTTGAAGTCGT-----AAGCAGTCGTGCTGAATGTTGGAAGA	3969
Qy	514	uLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla-----	531
Db	3968	AATCGTCCCTTGTCAAAATCGAAACAGTGTCCATTCAGTCGCTCACTCAGAACGGAC	3909
Qy	532	-----IleGl	533
Db	3908	AGGTGCTGAGTGGACGACGCTTGTCTACTCAATGTTGCTCAAGATGGACCAATGGC	3849
Qy	533	nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGl	553
Db	3848	TAAGAACCCTATGTCACCAACGACACAGGACAAG-----	3812
Qy	553	yLeuArgGluMetTyrProGlyArgAlaLeuProSer-----	565
Db	3811	---GTCGAATTCACCCACCTCGTTTCAACGATACCTTCCTCAATGATGGAANAATGT	3756
Qy	566	----AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGl	584
Db	3755	CCAGCAGTGGGTATCTCTCGTACGCTTGTGGGGTCAC-----CA	3714
Qy	584	nAspGlyAlaTrpTyrAsnTyrHisThrGlyLysGlyValLeuGluLysGlnAspPr	604
Db	3713	AATCCCTCGCTGGTACAAT-----GCTGATGGTGAAATGATGTCTCGCGAAGAAGCTCC	3660
Qy	604	oLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp	617
Db	3659	AGAAGTGCAGGATGGACTCAGGACGAGACGCTCTGGAT	3620
RESULT 11			
AE009858		14169 bp	DNA
LOCUS			linear
DEFINITION			BCT 16-JAN-2002
			Pyrobaculum aerophilum strain IM2 section 113 of 201 of the
			complete genome.
ACCESSION			AE009858 AE009441
VERSION			AE009858.1 GI:18160615
KEYWORDS			
SOURCE			Pyrobaculum aerophilum.
ORGANISM			Pyrobaculum aerophilum.
			Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;

REFERENCE		Thermoproteaceae; Pyrobaculum.
AUTHORS		1 (bases 1 to 14169) Fitz-Gibbon,S.T., Ladner,H., Kim,U.-J., Stetter,K.O., Simon,M.I. and Miller,J.H.
TITLE		Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum aerophilum
JOURNAL		PROC.Natl. Acad. Sci. U.S.A. 99 (2), 984-989 (2002)
PUBMED		11792869
AUTHORS		2 (bases 1 to 14169) Fitz-Gibbon,S.T., Ladner,H., Kim,U.-J., Stetter,K.O., Simon,M.I. and Miller,J.H.
TITLE		Direct Submission
JOURNAL		Submitted (12-DEC-2001) Microbiology and Molecular Genetics, University of California, Los Angeles, 405 Hilgard Ave, Los Angeles, CA 90095-1489, USA
FEATURES		Location/Qualifiers
source		1. 14169 /organism="Pyrobaculum aerophilum" /strain="IM2" /db_xref="taxon:13773" 144..1019 /gene="PAE2116" 144..1019 /gene="PAE2116" /note="Protein fate; Protein modification and repair" /codon_start=1 /transl_table=11 /product="methionine aminopeptidase" /protein_id="AAL63956.1" /db_xref="GI:18160616" /translation="MLRILRQVGVVNVKALKYALDLTPQDMPVLELCERVEAFTRAND AKPAPVNVSVNEVAHYATKRGDSLKPRTGLKIDVGAORDGYIVDAAVTVTLGPV FNNLKAQSALESAINAKPGIKAWQIGVEVRYIKSFGLSPIYNTLGHKVERYLLH AGHPVNPYSDKTSQALPAGDVVAIEPFGATNGEYVVDGEITTYRLARKHKNQYHL IDIISAEAGPLPTPRWFPOLTDDAIREALKAGVLQGYEVLVERSKGVAQFEDTVV GENEVIPLVSTLELL" 1060..2709 /gene="PAE2117" 1060..2709 /gene="PAE2117" /note="Protein fate; Protein folding and stabilization" /codon_start=1 /transl_table=11 /product="thermosome (chaperonin) alpha subunit" /protein_id="AAL63957.1" /db_xref="GI:18160617" /translation="MSQAVLTQIGGVPLVLKECTQRAFCKEALRLNIMIRATAEVM RTILPGKMDKMLIDSLCDITINDGATILDEMDYQHPIAKLLYEISKSQEEAGDGT TTAVLAGALLEAEAEKLEKRNHPYIVSGFKALDVAEEHLRKAIVPNRTDVTDLR KIATMSGGKISETVKEYFADLAVKAVLQVAEERNGKYVLDLDNIQIVKHHGSLDT QLVYGIYVDKEVHAAMPKRVNAKIALLDAPLEVEKEPEIDAEIRINDPTQMRALFE EERILRGYVDKLSLGVTLFTTKGIDDIQAYYLAAGILAVRRVKRSIDIEKLVRATG ARLVTSIEDTEADLPGALVEERRVGEKWFVEQCKNPKAVSILVRGGERLAVDEA ERNLDDALSVDVEDPYILPAGCAAEIEAAKAYRAFAPKVGREGQYAYEAFARALE NPKALAEAGLDPIDILTEUTHRHEQTDGKMGYIGLVYQKRVDMVSLGLVEPLTVKI complement(2706..3257) /gene="PAE2118" complement(2706..3257) /gene="PAE2118" /note="Hypothetical" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAL63958.1" /db_xref="GI:18160618" /translation="MLAEYVAQRVALIQERAYHRLVILSATSSVMSSGLALFFEP AFKYLALLALGLVLSRRVLNAAEKIRKFMANVEREAMARYFATLKEYINPIDL TKLYAETPIEDPIIAAEKKIRASTYILFSQSITRALRKNRLVITITALTISIPHLVN KDIIAIGVILLIVILLFKRSWV" complement(3247..4278) /gene="PAE2119" complement(3247..4278)



Db 9393 AAGCTGGGCAAAATACCTCCGCTCTCTCCCGCTGCATATGCCGTAGAGCGCGCAAG 9452  
Qy 187 -----PheLeuIleValTyrAsp 192  
Db 9453 ATGGGTGAACGTGCTAGAGAGGCGGCGCTTCGCGTGATTTTCGCTGTGTGTACGAT 9512  
Qy 193 -----ValPheLysAspValArgMetValPhe-----Ala 202  
Db 9513 GGGACTGTGTGTAGAGACGAATTTCTGCTTTTCATCGCGCGCCCTGAACACGCTGGAAGC 9572  
Qy 203 ProProSerSerValGlyLysPheGlyLysPheThrAspAsnTrpMetTrpProArgHis 222  
Db 9573 CCGATTATGACCCCGACCACTTC-----ACCCTTCTGCGTGTGGCTCTCGAAG 9623  
Qy 223 ThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgPProAlaGluTyr 242  
Db 9624 GAGCTTGGCCTTAGCCCAAGCTCAGTGTATAGGGC-----AGAGGGCGCCACCAT 9674  
Qy 243 SerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyr 262  
Db 9675 -----ATTGCCCTTCAAGGCGCGGCTGACCGCTTA 9707  
Qy 263 LysAlaAspTyrAlaMetThrIleGlyPheProGlySerThrAsp----- 278  
Db 9708 CTGGAG-----GCTGTAAACCGCGCGCTACCCGCTTGCACGAGCTTAGGGATGCC 9758  
Qy 279 -----ArgTyr 280  
Db 9759 CTGACAGAAATTCGCGCGCTTTAAAGCTGTTACCGCGGCGGCTTAAAGGAATTC 9818  
Qy 281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300  
Db 9819 GGCATTGACTGGGTACGACCTCAGAACCGAAATGTTTTCAAAAAGCTAGAGGAAGTT 9878  
Qy 301 ArgGlyIleLysGlnGly-----IleTrpLysGlu---AlaMetSerAlaAspGln----- 316  
Db 9879 GTTACAATGGCAGAGGACTACGTCTACAGAAACGCTCAGCTGAGAGGGATCAGCTGGAC 9938  
Qy 317 -----AlaThrArgIleLysTyr-----AlaSerLysTyrAla 327  
Db 9939 ACAAGCGGACAGTTCACCAAGACTGTATCCGCTTTAAGCTTGGCGGGAGGAGTGCGG 9998  
Qy 328 GlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAsp 347  
Db 9999 TACATAACCGTATACCTGACCTGGCAGT-----AAGCTACTAGCCCAATTCGAC 10046  
Qy 348 ValIleGlyArgLysArgAlaGluGluArgAlaPhe----- 359  
Db 10047 GGC---TCTCAGAAAAGCGCAAGTATAGCCTCTATCATTCAGCGCTTTGGCGGCAAG 10103  
Qy 360 AlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeu 379  
Db 10104 GCGAGGTCAACCAACGAATGCCAATGGACTGTTAC----- 10142  
Qy 380 GluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSer----- 397  
Db 10143 -----CTCCACACCGGCGGCAATATGTCCTATTCGCCATGACAGCTGTTAAATGCTGTG 10196  
Qy 398 -----GluThrLeuPhe-----GlyGlyThrGluValValArgPheAlaGln 411  
Db 10197 AAGGCGTTTGTGACGAACCTTACAGTAAGGGGTGATCAGCGAGGATAGATAGACACAG 10256  
Qy 412 PheAlaAsnAlaLeuAlaThrAspProAspAla-----HisAlaGlyIleLeuLysSer 429  
Db 10257 TTGTTGAGGACATAGTACAGGCGCCCAACACCGCTCACTATGCGGGGTGGAGTTCTCG 10316  
Qy 430 LeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMet 449  
Db 10317 GTTTATCACAAGACGAAGGTAATACGATTTTATAGAAATAAGGTATCACCCAGGC--- 10373  
Qy 450 LeuAspIleValArgArgIleProAlaAspLysLeuProAspIlePheLysAsnVal 469  
:::|||||

Db 10374 -----AACGAGGCTCTAAAAATGCC 10394  
Qy 470 IleAspLysLysPheLysGly-----AspThrLysLys 480  
Db 10395 GTTCTGAACGCCCTAAAGCAAGAGGCTGAAAGAGGTTGACACTTCACCTGCTGAAGAG 10454  
Qy 481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500  
Db 10455 TACGGCAGCTACGAATTCACGTGGCTGTGAAATCGTATACAAAAGCTTGTGAAGACCTTG 10514  
Qy 501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520  
Db 10515 GCAGTAGCTGGCTGAGAAAGGAGAGTACTACGCCATTGAT----- 10556  
Qy 521 LeuSerLysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540  
Db 10557 ---CACAGAAAGCGTGTAAATCAGCGTCAAGCGGAGCACAAGACGAGTTGTAATACC 10613  
Qy 541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560  
Db 10614 -----CTAAAACCGCTGGCTGGAGGAG----- 10637  
Qy 561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580  
Db 10638 -----GATAGACACTTTACGGTGAAGTGG----- 10661  
Qy 581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600  
Db 10662 -----TACTAGTAAATCGTATTACATATGAAGGCTCCGCGAA 10706  
Qy 601 LysGln-----AspProLysSerAspGluPheAlaValGlnGluAsn 614  
Db 10707 ATCCAGCGTATGGCGCTAAACCGCGACGTGAGGCTGAGCGCTTTATA-----AGGCG 10760  
Qy 615 IleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHis 634  
Db 10761 CTTGAGGATGTGCTGAGG---CGCCGCCACGCCCAACACGAGTGAAGGCTGATTGAA 10817  
Qy 635 IleAlaPheLeuSerAsnAsnAspIleThrGlyLysAsnSerGlySerPro-----Val 652  
Db 10818 GTGTTGACCGCGCGGAGGAGGAG-----GGGACGATAGATTACCGCTAGAAGTG 10868  
Qy 653 PheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSer 672  
Db 10869 CGCCACGAAAGGCGCAACGTGTCGCGCGCTGTGGACTGTGAGTATGAATCTGTAATG 10928  
Qy 673 GlyAsp-----IleGlu 676  
Db 10929 GCGCATCAACCCGTTGGCCAGTGGCGGGGAGGAGCTGCCCGCTGCCATATAGTGGAG 10988  
Qy 677 PheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIle 696  
Db 10989 TACGACAGAGGCGGAAAGGAGA-----CAGCTAAAGATAGTATGTTACTGGGTT 11039  
Qy 697 AspLys 698  
Db 11040 AAGAAA 11045

## RESULT 12

A81135

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Sequence 1 from Patent EP0909818.

A81135.1 GI:6731609

unidentified.

unclassified.

1 (bases 1 to 2652)

Brown, J.R. and Wang, M.

Valyl tRNA synthetase (EC 6.1.1.9) from Streptococcus pneumoniae

Patent: EP 0909818-A 1 21-APR-1999;

SMITHKLINE BEECHAM CORP (US)

2652 bp DNA linear PAT 21-JAN-2000

FEATURES		Location/Qualifiers		BASE COUNT		ORIGIN	
source		1. 2652		733 a		612 c 621 g 686 t	
		/db_xref="taxon:32644"					
		/organism="unidentified"					
Alignment Scores:		5.67		Length:		2652	
Pred. No.:		130.00		Matches:		106	
Score:		32.81%		Conservative:		62	
Best Local Similarity:		20.70%		Mismatches:		166	
Query Match:		3.50%		Indels:		178	
DB:		6		Gaps:		25	
US-10-008-355-2 (1-712) x A81135 (1-2652)							
Qy	192	AspValPhe	-----LysAspValArgMetValPheAlaPro	203			
Db	76	GATGTTTTCAGCCCTTCAGCGCATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA	135				
Qy	204	ProSerSerValGlyLysPheGlyAspThrAspAsnTrpMetTrpProArgHisThr	223				
Db	136	CCAAACGTTACAGGTAACTT	-----CACCTT	162			
Qy	224	GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer	243				
Db	163	GGT	-----CACGCT	171			
Qy	244	LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys	263				
Db	172	TGGGATACA	---ACTTTGCAAGATATTATCATCCGTCGCAAAACGCATGCAAGGTTTGTAT	228			
Qy	264	AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu	---Thr	282			
Db	229	-----ACCCTTTGGCTTCCTGGGATGGACCAACGAGGATGGCCACT	270				
Qy	283	SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly	302				
Db	271	CAGGCTAAGGTAGAGGAGCGCTTCGCTGGTGAGGCAATTTCCCGCTATGACCTTGGTCT	330				
Qy	303	-----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr	318				
Db	331	GAGTCTTTTGTGACGAAGTCTGGGAATGGAAGAC	-----GAATATGCCACT	378			
Qy	319	ArgIleLysTyrAla	---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle	337			
Db	379	ACTATCAAGGAACAAATGGGCAAGATGGGGCTCTCTGAGACTATTCTCGTGAGCGTTTC	438				
Qy	338	GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArg	357				
Db	439	ACTCTTCAGCAAGGTTGTCA	-----AAAGCTGTTCTGTAAG	474			
Qy	358	AlaPheAlaAspTrpIleArgLys	-----AsnGlyLysSerAlaValTyrGly	373			
Db	475	GTCTTTGTGACCTTTTACAAAGAAAGCTGGATCTACCGTGGTGGTATCATCAACTGG	534				
Qy	374	Asp	-----ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys	388			
Db	535	GACCCACAGCTCGCACAGCCCTTCTGATATTGAGGTGATTCACAGGATGTGGAAGGT	594				
Qy	389	AlaAsnArgGluMetThrTyrLeu	-----	396			
Db	595	GCCTTCTACCATGATTAATCATCTCGAAGATGGTTTCACCGGTCCTTGAAGTTGCTACA	654				
Qy	397	-----SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn	414				
Db	655	ACTCGCTCTGAGACCATGTTTGGGGACGTT	-----	684			
Qy	415	AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr	434				
Db	685	CGGTTTGGCGTCAACCCAGAA	-----GACCCCGCGTAC	717			
Qy	435	LysAsp	-----TyrLeuProSerLeuAspArgLysValLeuProAla	448			
Db	718	AAGGACTTGATTGGTAAATAATGTCATCTTCCA	---ATCGCTAATAAATCATCCCAATC	774			
Qy	449	MetLeuAsp	-----IleValArgArg	455			
Db	775	GTTGGAGATGAGCAGCAGATCCTGAGTTTGGTACTGGTTCGTTGAATAATCACACCTGCC	834				
Qy	456	ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys	475				
Db	835	CACGATCCAAATGACTTCTTGGTGGCCACAGTCATAACTTGGCCACAGTCAACGCTCATG	894				
Qy	476	GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp	495				
Db	895	AACGACGACGGAACCATGAAATGACTTGGCCTTTGAATTTTTCAGGCATG	-----GAC	945			
Qy	496	LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys	515				
Db	946	CGTTTTCGAAGCTCGT	-----AAGCAGCTCGTTGCTAAGTTTGGAAAGAAATC	990			
Qy	516	AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla	-----	531			
Db	991	GGTGCCCTCGTCAAAATCGAAACGTTGCCACAGTGTGGTCTACTCAGAGCGTACAGGT	1050				
Qy	532	-----	-----IleGlnAla	534			
Db	1051	GTTGTGTTGAACCTCGCTTGTCTACTCAATGTTCTGTCGAAGATGGACCAATTGGCTAAG	1110				
Qy	535	AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeu	554				
Db	1111	AAGCCATTGCCAACCAACACAGACAGAGGACAAG	-----	1143			
Qy	555	ArgGluMetTyrProGlyArgAlaLeuProSer	-----	565			
Db	1144	GTCAATTCACCCCTCGTTTCAACGATACCTTCTCAATGGATGGAAATGTCCAC	1203				
Qy	566	AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp	585				
Db	1204	GACTGGGTTATCTCTCGTCAGCTCTGGTGGGTCTAC	-----CAAAATC	1245			
Qy	586	GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys	605				
Db	1246	CCTGCTCGGTGACAAAT	-----GCTGATGGTGAATGATGTGCGGCGAAGAGCTCCAGAA	1299			
Qy	606	SerAspGluPheAlaValGlnGluAsnIleLeuAsp	617				
Db	1300	GGTGACGGATGGACTCAGGACGAAGACGTCTTGAT	1335				
RESULT 13							
LOCUS		AR065782		2652 bp		DNA linear	
DEFINITION		Sequence 1 from patent US 5849555.				PAT 29-SEP-1999	
ACCESSION		AR065782					
VERSION		AR065782.1		GI:5995998			
KEYWORDS		Unknown.					
SOURCE		Unknown.					
ORGANISM		Unknown.					
REFERENCE		1 (bases 1 to 2652)					
AUTHORS		Brown, J.R., Jaworski, D.D., Lawlor, E.J. and Wang, M.					
TITLE		Valyl trna polynucleotides from Streptococcus pneumoniae					
JOURNAL		Patent: US 5849555-A 1 15-DEC-1998;					
FEATURES		Location/Qualifiers					
source		1. 2652		/organism="unknown"			
BASE COUNT		733 a		612 c		621 g 686 t	
ORIGIN							
Alignment Scores:		5.67		Length:		2652	
Pred. No.:		130.00		Matches:		106	
Score:		32.81%		Conservative:		62	
Best Local Similarity:		20.70%		Mismatches:		166	
Query Match:		3.50%		Indels:		178	
DB:		6		Gaps:		25	

Query Match: 3.50% Indels: 178  
DB: 6 Gaps: 25  
US-10-008-355-2 (1-712) x AR065782 (1-2652)

QY 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203  
DB 76 GATGTTTTCAGGCTTCAGCGATCAAAAGCGTATTATTCATTCGTTATTCACCA 135  
QY 204 ProSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223  
DB 136 CCAAACGTTACAGTAACCTT-----CACCTT 162  
QY 224 GlyAspPheSerValPheArgValTyAlaGlyAlaAspAsnArgProAlaGluTySer 243  
DB 163 GGT-----CAGGCT 171  
QY 244 LysAspAsnLysProTyLysProValTyPheAlaAlaValSerMetGlnGlyTyLys 263  
DB 172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAACGCATGCAAGGTTTGTAT 228  
QY 264 AlaAspAspTyAlaMetThrIleGlyPheProGlySerThrAspArgTyLysLeu---Thr 282  
DB 229 -----ACCTTTGGCTTCTCGGGATGACACCGCAGGAGTGCCTACT 270  
QY 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302  
DB 271 CAGGCTAAGGTAGAGGAGCGCTTCGCTGGTGCAGGCGATTTCCCGCTATGACCTTGGTCTG 330  
QY 303 -----lleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318  
DB 331 GAGTCTTTCTTGACGAAGTCTGGGATGAAAGAC-----GAATATGCCACT 378  
QY 319 ArgIleLysTyAla---SerLysTyAlaGlnSerAlaAsnTyTrpLysAsnSerIle 337  
DB 379 ACTATCAAGAACAAATGGCGCAGATGGGCTCTCTAGACTATTCTCGTAGCGGTTTC 438  
QY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArg 357  
DB 439 ACTCTTGACGAAGTTGTCA-----AAAGCTGTTGCGTAAG 474  
QY 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyGly 373  
DB 475 GTCTTTGTGACCTTTACAGAAGGCTGATCTACCGCTGGTGAGTTATCATCACTACCTGG 534  
QY 374 Asp-----ValLeuSerSerLeuGluLysAlaTyLysGluGlyAlaLys 388  
DB 535 GACCCACGACTCGCACAGCCCTTCTGATATTGAGGTGATTCACAGGATGCGAAGGT 594  
QY 389 AlaAsnArgGluMetThrTyLysLeu----- 396  
DB 595 GCCTTCTACCATGATGAATTACATGCTGGAAGATGTTTCACGCGCTTGAAGTTGCTACA 654  
QY 397 -----SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414  
DB 655 ACTCGTCTGAGACCATGTTTGGGACGTT----- 684  
QY 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTy 434  
DB 685 GCGGTTCCGCTCAACCCAGAA-----GACCCCGGCTAC 717  
QY 435 LysAsp-----TyLeuProSerLeuAspArgTyLysValLeuProAla 448  
DB 718 AAGGACTTGTATGGTAAATAATGTCATCTTCCA---ATCGCTAATAAATCATCCCAATC 774  
QY 449 MetLeuAsp----- 455  
DB 775 GTTGGAGATGAGCAGCAGATCTCGAGTTGGTACTGGTCTGCGTGAATAATCAGACCTGCC 834  
QY 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475  
DB 835 CAGCATCAAATGACTTCTTGGTTGGCCACGTCATATCTTGCACAAAGTCACAGTCAATG 894

QY 476 GlyAspThrLysLysTyAlaAspPheValPheAspLysSerValValProTySerAsp 495  
DB 895 AACGACGACGGAACCATGATGACTTGCCTTTCGAATTTTCAGGCATG-----GAC 945  
QY 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515  
DB 946 CGTTTTCGAAGCTCGT-----AAGCAGCTCGTTGCTAAGTTTGGAAAGAAATC 990  
QY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531  
DB 991 GGTGCCCTCTCGTAATTCGAAAACGTCCTCACAGTGTGGTCTACAGACGCTACAGGT 1050  
QY 532 -----IleGlnAla 534  
DB 1051 GTTGTGTTGAACCTCGCTTGTCTACTCAATGTTGTCGAAGATGGACCAATTCGCTAAG 1110  
QY 535 AspAlaMetAlaAsnAlaTyAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554  
DB 1111 AAGCCCATTTGCCAACCAAGACACAGAGGACAAG----- 1143  
QY 555 ArgGluMetTyTrpGlyArgAlaLeuProSer----- 565  
DB 1144 GTCGAATTCATCCACCTCGTTTCAACGATACCTTCTCAATGATGGAAATGTCCAC 1203  
QY 566 AspAlaAsnPheThrMetArgMetSerTyTrpClySerIleLysGlyTyTrpGluProGlnAsp 585  
DB 1204 GACTGGGTTATCTCTCGTACGCTGCTGGTGGGTCCAC-----CAAATC 1245  
QY 586 GlyAlaTrpTyTrpAsnTyTrpHisThrGlyLysGlyValLeuGluLysGlnAspProLys 605  
DB 1246 CCGTCCCTGGTACAAAT-----GCTGATGTCGTAATGATGTCGGCGAAGAGAGCTCCAGAA 1299  
QY 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617  
DB 1300 GTGACGACGAGGACTCAGGACGAGAGAGCTCTTGGAT 1335

RESULT 14  
E26074  
LOCUS Novel vals. 2652 bp DNA linear PAT 18-JUN-2001  
DEFINITION Novel vals.  
ACCESSION E26074  
VERSION E26074.1 GI:13025049  
KEYWORDS JP 1999113577-A/1.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 2652)  
AUTHORS James,R.B., Deborah,D.J., Elizabeth,J.L. and Ming,W.  
TITLE Novel vals  
JOURNAL Patent: JP 1999113577-A 1 27-APR-1999;  
SMITHKLINE BEECHAM CORP  
COMMENT OS Unidentified  
PN JP 1999113577-A/1  
PD 27-APR-1999  
PF 17-OCT-1997 JP 1997321886  
PI JAMES R BURAU,DEBORAH D JAWORUSUKI,ELIZABETH JANE LAWLOR, PI  
PC C12N15/09,A61K35/76,A61K38/43,A61K39/395,A61K48/00,C07K14/315,  
C07K14/335,  
PC C07K16/12,C12N5/10,C12N9/00,C12P21/02,C12Q1/00,G01N33/566, PC  
G01N33/569,  
PC C12N15/00,A61K37/48,C12N5/00  
CC Strandedness: Single;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..2652 /organism='Unidentified'.  
FT Location/Qualifiers  
1..2652 /organism='unidentified'  
/db\_xref='taxon:32644'  
BASE COUNT 733 a 612 c 621 g 686 t

## ORIGIN

## Alignment Scores:

Pred. No.: 5.67 Length: 2652  
 Score: 130.00 Matches: 106  
 Percent Similarity: 32.81% Conservative: 62  
 Best Local Similarity: 20.70% Mismatches: 166  
 Query Match: 3.50% Indels: 178  
 DB: 6 Gaps: 25

US-10-008-355-2 (1-712) x E26074 (1-2652)

Qy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203  
 Db 76 GATGTTTCAAGCGCTTCAGCGGATCAAAAGGCTAAGCGCTTATTCAATCGTTATTCCACCA 135  
 Qy 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223  
 Db 136 CCAACGGTTACAGGTAAACTT-----CACCTT 162  
 Qy 224 GlyAspPheSerValPheArgValTyraGlyAlaAspAsnArgProAlaGluTySer 243  
 Db 163 GGT-----CACGCT 171  
 Qy 244 LysAspAsnLysProTyTyLysProValTyPheAlaAlaValSerMetGlnGlyTyLys 263  
 Db 172 TCGGATACA---ACTTTGCAAGATATTATCATCGCTCAAAACGCGATCAAGGTTTGTAT 228  
 Qy 264 AlaAspAspTyAlaMetThrIleGlyPheProGlySerThrAspArgTyLeu---Thr 282  
 Db 229 -----ACCTTTGGCTTCTGGGATGGACCGACGAGGATGGCCACT 270  
 Qy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302  
 Db 271 CAGGCTAAGGTAGAGGACGCTGCGTGGTCAGGCGATTTCCCGCTATGACCTTGGTGT 330  
 Qy 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318  
 Db 331 GAGTCTTTTTCAGCAAGATCTGGGAATGGAAAGC-----GAATATGCCACT 378  
 Qy 319 ArgIleLysTyAla---SerLysTyAlaGlnSerAlaAsnTyTrpLysAsnSerIle 337  
 Db 379 ACTATCAAGGAACAAATGGGGCAAGATGGGCTCTGTAGACTATTCTCGTAGCGGTTTC 438  
 Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluLuar 357  
 Db 439 ACTCTTGACGAAGTTTGCTCA-----AAAGCTTTCGTAAAG 474  
 Qy 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrgly 373  
 Db 475 GTCTTTGTGGACCTTTACAAGAAAGGCTGGATCTACCGTGGTGGATTTATCATCAACTGG 534  
 Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTyLysGluGlyAlaLys 388  
 Db 535 GACCCAGCAGCTCGCACAGCCCTTCTGATATTGAGTGATTCACAGGATGTGGAAGGT 594  
 Qy 389 AlaAsnArgGluMetThrTyLeu----- 396  
 Db 595 GCCTTCTACCATCAATTAATCATGCTGGAAGATGTTCCACCGCTCTTGAAGTTGCTACA 654  
 Qy 397 -----SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414  
 Db 655 ACTCGTCTGAGACCATGTTTGGGACGCTT----- 684  
 Qy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTy 434  
 Db 685 CGGTTGCGGTCAACCCAGAA-----GACCCGCGCTAC 717  
 Qy 435 LysAsp-----TyLeuProSerLeuAspArgLysValLeuProAla 448  
 Db 718 AAGGACTTGATGGTAAAAATGTATCTCTTCCA---ATCGCTAATAAATCATCCCAATC 774  
 Qy 449 MetLeuAsp-----IleValArgArg 455

Db 775 GTTGAGATGAGCACGCGAGATCCTCAGTTTGTTACTGTTGCTGTAATAATCACACCTGCC 834  
 Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475  
 Db 835 CACGATCCAAATGACTTCTTGGTTGGCCACGCTCATAACTTGGCCACAGTCAAGTCAATG 894  
 Qy 476 GlyAspThrLysLysTyAlaAspPheValPheAspLysSerValProTySerAsp 495  
 Db 895 AACGACGACGGAACCATGATGACTTGGCCTTTGAATTTTTCAGGCATG-----GAC 945  
 Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515  
 Db 946 CGTTTGAAGCTCGT-----AAGCAGCTCGTTGCTAAGTTGGAAGAAATC 990  
 Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531  
 Db 991 GGTGCCCTCGTCAAAATCGAAAAACGCTCCACAGTGTGTCCTCACTCAGACGCTACAGGT 1050  
 Qy 532 -----IleGlnAla 534  
 Db 1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGTTTCGTCAAGATGGACCAATTTGCTAAG 1110  
 Qy 535 AspAlaMetAlaAsnAlaTyAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554  
 Db 1111 AACGCCATTGGCAACCAAGACACAGAGGACAAG----- 1143  
 Qy 555 ArgGluMetTyProGlyArgAlaLeuProSer----- 565  
 Db 1144 GTCGAATTCACCCACCTCGTTTCAACGATACCTTCTTCAATCGATGGAAAAATGTCCAC 1203  
 Qy 566 AspAlaAsnPheThrMetArgMetSerTyrglySerIleLysGlyTyrglyProGlnAsp 585  
 Db 1204 GACTGGGTATCTCTCGTCAGCTCTGGTGGGTCAC-----CAATC 1245  
 Qy 586 GlyAlaTrpTyArgAsnTyHisThrGlyLysGlyValLeuGluLysGlnAspProLys 605  
 Db 1246 CTTCCCTGGTACAAT-----GCTGATGGTGAATGTATGTCGCGGAAGAGCTCCAGAA 1299  
 Qy 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617  
 Db 1300 GGTGACGGATGGATCAGGACGAAGACGTCCTTGGAT 1335

RESULT 15  
 LOCUS AE013500  
 DEFINITION Methanosarcina mazei strain Goel, section 282 of 379 of the complete genome.  
 ACCESSION AE013500 AE008384  
 VERSION AE013500.1 GI:20907162  
 KEYWORDS  
 SOURCE  
 ORGANISM Methanosarcina mazei Goel.  
 Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 Methanosarcinaceae; Methanosarcina.  
 1 (bases 1 to 9827)  
 Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,  
 Martinez-Arias,R., Henne,A., Wierzer,A., Baumer,S., Jacobi,C.,  
 Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,  
 Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,  
 Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.  
 The genome of Methanosarcina mazei: Evidence for lateral gene transfer between Bacteria and Archaea  
 J. Mol. Microbiol. Biotechnol. (2002) In press  
 2 (bases 1 to 9827)  
 Deppenmeier,U., Johann,A., Hartsch,T., Merkl,R., Schmitz,R.A.,  
 Martinez-Arias,R., Henne,A., Wierzer,A., Baumer,S., Jacobi,C.,  
 Brueggemann,H., Lienard,T., Christmann,A., Boemecke,M., Steckel,S.,  
 Bhattacharyya,A., Lykidis,A., Overbeek,R., Klenk,H.-P.,  
 Gunsalus,R.P., Fritz,H.-J. and Gottschalk,G.  
 Direct Submission  
 Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute for Microbiology and Genetics, Grisebachstrasse 8, Goettingen

[illegible]



Qy	59	LysProSerIleAlaAsnAlaValValIlePheGlyGlyCysThrGlyIleThrVal	78
Db	3192	AAACCCGGATCAGGAA-----ATCCAGCTCGGGGATCTCTCCGAAGGATTCAGAAA	3245
Qy	79	SerAspGlnClyLeuIlePheThrAsnHisHisCysGlyTyr-----	92
Db	3246	GCTTTTTCGAAATCGTTGATGCAGCCAGCAATTCGTAAGTTTAAAGACTGCACCCACCG	3305
Qy	93	-----GlyAlaIleGlnSerGlnSer	99
Db	3306	GACGACCGGGCTGTCGGCTACTGAAGCTGTTAAGGATGGGATATTCTCTGAGGAACGG	3365
Qy	100	ThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeu	119
Db	3366	CTTGCCAGCTATCACAGGTTAACTGACGAACCTGCGTTTCAGTCAAGAAGATCAGAAATC	3425
Qy	120	ProIleProGlyLeuSerValIysTyrLeuArgLysIle-----ValLysValThr	136
Db	3426	GGGTTGAAGAGGCTTCAAGAAAGAAAGTTTCAAGAAATAGCAGTGGACATTAATAAATAT	3485
Qy	137	AspLysValGluGlyGlnLeu-----LysGlyIleThr	147
Db	3486	AGAAATATCCACCGGAGAGCTGAAGATAGGTTTATTTTAAAGATGCAAGCATATTCAG	3545
Qy	148	AspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGlu	167
Db	3546	GGTTTATCAATCGTTTTCGACCTGGGAGGAACATC-----	3581
Qy	168	AsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyrPhe	187
Db	3582	-----GGAGGAGTAGTTTACATTGAACCTCCCTGACTTACAAT-----	3620
Qy	188	LeuIleValTyrAspValPheLysAspValArgMetValPheAlaProProSerSerVal	207
Db	3621	-----TACCCTGACCTACAAATTTATATGATAAAAAAACCCTATGCTATT	3662
Qy	208	GlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSer	227
Db	3663	TCT-----ATGTCACGAAGACTACAGGAT---TCT	3689
Qy	228	ValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLys	247
Db	3690	GCTTTTCGTTATTG-----AAGAAGACAGCCTACGATATAGATCATGAT	3734
Qy	248	ProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyr	267
Db	3735	AGATATAAAAAAGCCCTT-----	3752
Qy	268	AlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGlu	287
Db	3753	-----GAAACCTTGATAGGTT-----	3770
Qy	288	AspArgIleGluAsnGlu---AsnAsnProArgIleGluValArgGlyIleLysGlnGly	306
Db	3771	GAAGAAGTCAAAAAAGGAAGTGAATAACCTGAGATTTCATAT-----	3812
Qy	307	IleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyr	326
Db	3813	-----CATCTCTCGTTTTTTTAAAGGTTTTCGAAATATC	3845
Qy	327	-----AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg	341
Db	3846	AGGATTCAGTCAACTCGAAGAGTCGCTTCGACCTTTTCGGAAGGCCCTGGGAATTTAGCTGG	3905
Qy	342	GlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAsp	361
Db	3906	GACCTTTT-----TCAAAAGAACCGGAAAAAAGAGATTACCGGTCT	3947
Qy	362	TrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLys	381
Db	3948	TTTATCGGAGACAGATTGGAAGT-----ATTGGAGATATCTTCTGAAGCTTGAGAT	4001

Qy	392	AlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPhe	401
Db	4002	CCCCGTTAAAGCA	4046
Qy	402	GlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAsp	421
Db	4047	GAATAAACG--GTTCGGGTGTTGAGAAGTTATTGAGAGTCAGATCCCGAAATATCCGGAG	4103
Qy	422	-----AlaHisAlaGlyIleLeuLysSerLeuAspAsp	432
Db	4104	TACCTGGAAGATTTCCTGAAACGATCGAACACATAGGGCTCTGTTTGAAGCCGCGAGAA	4163
Qy	433	LysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIle	452
Db	4164	TTGACCGGAGACGACGATTCCACATTGTTGATAAAAAA-----TTTCGATCTC	4208
Qy	453	ValArgArgIleProAla-----	459
Db	4209	ATTGAGAAATGTCTCAAAAGTGAACCTGAAATATCCGAACACCTCAGCAGCGCTTGATGAT	4268
Qy	459	-----	459
Db	4269	TCTCTCATGAATTTTGAAGGATCTCTGAGGAGAGGAGAAATATTAGAGAGGAGGATATATA	4328
Qy	460	-----AspLysLeuProAspIlePheLysAsnValIleAspLys	472
Db	4329	GAAGAGCAAAAGCGCGTTTATGATGGGCAATAGAGATTTCAGGAAAGTCTCTCGATAAA	4388
Qy	473	--LysPheLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValVal	491
Db	4389	GACCTCGAAAACTCCGAAACCAAGTTTACCTGAGTTACATATACCGTTACATGGCAGGT	4448
Qy	492	ProTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAla--	510
Db	4449	ATGTATTTCGGACACG--GAAATCTCGTAAAGGCAGAGGAATGTTATAAAGACGATTC	4505
Qy	511	-----LysAlaIleGluLysAspProAla-----Val	519
Db	4506	GGCCTGCTTGAACCGGAGATCGAAATGAGCCTGAAATATTCCGTTTCTGAAATAATT	4565
Qy	520	GluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsn	539
Db	4566	GAAGTTTCCGAAACATATACGTTCCCAAAAATATTCTTTTCTATGGCCATTGTCAGAT	4625
Qy	540	AlaTyrAlaIleGluLysGly-----	546
Db	4626	ATGTACAAA--GATTACGCGTTAATTTTCAGATACGGAAGATACCAAAAACCCAGG	4682
Qy	547	-----LysArgLeuPheAlaGlyLeuArgGluMetTyrProGlyArg	561
Db	4693	GACTACTATACGAAGCGGAGGAAACTTTCCGGGACCTGATTAATAATACCCGGAT--	4739
Qy	562	AlaLeuProSerAspAlaAsnPheThrMetArgMetSer-----TyrGly	576
Db	4740	CTGCTTGAAGCCGATAAAAACCTTCGCAGCTCTTTAGATGAACGTGCGAGAGCTGTTTCA	4799
Qy	577	SerIleLysGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHisThrThrGlyLys	596
Db	4800	GAGATCGGAGACTTCGAAAGTGCAGAAATGTCTATAAGGACACGACGATATATCATGAA	4859
Qy	597	GlyValLeuGluLys-----GlnAspProLysSerAspGluPheAlaValGlnGluAsn	614
Db	4860	AACTGTGCAAAAGAACCTGAAACCTGAAACCTGAAATATTCTCTTAAG-----	4913
Qy	615	IleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHis	634
Db	4914	CTTTCGCACATCTACCG-----GAGCTTGGTGACCTT--	4946
Qy	635	IleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAsp	654
Db	4947	-----TTTCGGAGAGGAGAAAATTT-----GAAAGCAAGCGCTTATTGAT	4991
Qy	655	LysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAsp	674



```
Db 4992 AAA-----GAA 4997
    |||
Qy 675 IleGluPheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPhe 694
    |||||
Db 4998 ATCGAAATTTACGAAATCTTCACCCACGACAGCCGATAAACTCATGTTAGAGCTAAT 5057
    |||||
Qy 695 MetIleAspLysTrpGlyGlnCysProArgLeuIleGlnGlu 708
    |||
Db 5058 AAAGCCGATACCTGGAACACAGATTGGAAGGCTGTATGCAGAA 5099
    |||
```

Search completed: May 23, 2003, 11:33:15  
Job time : 4235.8 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 23, 2003, 07:32:23 ; Search time 316.182 Seconds  
(without alignments)  
5071.208 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MQMKLKSTLLGALLLGASG.....LFMIDKWGQCPRLIGELKLI 712

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlh  
-O=/Cnp2\_1/USPTO.spool/US10008355/runat\_16052003\_110359\_2525/app\_query.fasta\_1.1230  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US10008355.ecgn\_1.1.275.0runat\_16052003\_110359\_2525 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3719	100.0	2139	24	AA43635	Porphyromonas ging
2	146.5	3.9	4104	24	ABN68149	Streptococcus poly
3	134.5	3.6	2853	20	AAK91754	Porphyromonas ging
4	134.5	3.6	2859	20	AAK91621	Porphyromonas ging
5	133	3.6	2852	23	AAK55847	Streptococcus pneu
6	130	3.5	2852	20	AAK57027	S. pneumoniae valis
7	130	3.5	2852	20	AAK32814	S. pneumoniae valis
8	130	3.5	2652	21	AAK14369	DNA encoding strep
9	130	3.5	2652	21	AAK14369	DNA encoding strep
10	129	3.5	6004	19	AAV52160	Streptococcus pneu
11	127	3.4	2652	23	AAK55949	Streptococcus pneu
12	126.5	3.4	4110	24	ABN68148	Streptococcus poly
13	123.5	3.3	2142	20	AAK31948	M. catarrhalis str
14	122.5	3.3	4716	24	ABQ69595	Listeria innocua D
15	122.5	3.3	4731	24	ABQ67873	Listeria innocua D
16	122.5	3.3	495269	24	ABQ67195	Listeria innocua c
17	121.5	3.3	2649	24	ABN66404	Streptococcus poly
18	121.5	3.3	2155361	24	ABN71527	Streptococcus poly
19	121	3.3	2646	24	ABN66405	Streptococcus poly
20	119.5	3.2	3271	22	AAI64839	XRN-100 coding seq
21	117.5	3.2	1664976	19	AAV21209	Methanococcus jann
22	117	3.1	1830121	17	AAK2063	Haemophilus influe
23	116.5	3.1	15424	21	AAK81763	Porcine Lelystad v
24	116.5	3.1	15424	21	AAK81764	Porcine Lelystad v
25	116	3.1	3598	21	AAK77776	Human cancer assoc
26	115.5	3.1	6822	24	ABN71008	Streptococcus poly
27	115.5	3.1	6834	24	ABN68971	Streptococcus poly
28	115	3.1	4853	17	AAK13139	Alpha-D-glucosyltr
29	114.5	3.1	1974	24	AAK46803	H influenzae BVH-N
30	114	3.1	2136	17	AAK11244	Neisseria meningit
31	114	3.1	3420	22	AAH14452	Human cDNA sequenc
32	114	3.1	3431	21	AAK76951	Human OREF ORF2506
33	113.5	3.1	2856	23	AAK52731	E. coli DNA for ce
34	113	3.0	8930	19	AAV22834	Haemophilus para
35	112.5	3.0	36471	21	AAK81453	N. meningitidis pa
36	112.5	3.0	349980	21	AAK21611	Neisseria meningit
37	112.5	3.0	837096	21	AAK81489	N. meningitidis pa
38	112	3.0	2175	23	AAK79571	DNA encoding novel
39	112	3.0	2175	23	AAK81718	DNA encoding novel
40	112	3.0	3285	22	AAH54398	S. epidermidis gen
41	112	3.0	56506	21	AAK69168	Bacteriophage Dp-1
42	111	3.0	4411529	22	AAI99682	Mycobacterium tube
43	110.5	3.0	3279	14	AAQ51556	Sequence encoding
44	110.5	3.0	12665	19	AAV52267	Streptococcus pneu
45	110.5	3.0	15450	21	AAK27809	North American por

ALIGNMENTS

RESULT 1

AA43635

ID AAL43635 standard; DNA; 2139 BP.

XX AAL43635;

AC AAL43635;

XX 05-SEP-2002 (first entry)

DT Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.

DE Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;  
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;  
KW periodontitis.  
XX Porphyromonas gingivalis.  
OS Porphyromonas gingivalis.

XX Key Location/Qualifiers  
FH CDS 1..2139

FT /\*tag= a  
 FT /product= "Porphyromonas gingivalis DPP-7"  
 PN  
 PN WO200238742-A2.  
 PN  
 PD 16-MAY-2002.  
 XX  
 XX  
 XX 08-NOV-2001; 2001WO-US46782.  
 XX  
 XX 08-NOV-2000; 2000US-246827P.  
 XX  
 XX (UYGE-) UNIV GEORGIA RES FOUND INC.  
 PA  
 XX Travis J, Potempa JS, Banbula A, Bugno M;  
 XX  
 XX WPI: 2002-490075/52.  
 XX  
 XX P-PSDB; AAO15205.  
 DR  
 XX

XX Novel isolated dipeptidylpeptidase useful for identifying inhibitor of  
 PT the dipeptidylpeptidase for protecting an animal from periodontal  
 PT disease caused by Porphyromonas gingivalis -  
 XX  
 XX Claim 11; Fig 4; 65pp; English.

XX The invention comprises the amino acid and coding sequence of the  
 CC Porphyromonas gingivalis dipeptidylpeptidase-7 (Dpp-7) enzyme. The Dpp-7  
 CC enzyme has amidolytic activity for cleavage of a peptide bond between the  
 CC second and third amino acids from the N-terminal end of a target peptide.  
 CC The Dpp-7 target peptide has an aliphatic or aromatic residue as a  
 CC substituent on the alpha-carbon atom of the second amino acid from the  
 CC N-terminal end. The Dpp-1 protein and DNA sequences of the invention are  
 CC useful for identifying a Dpp-7 inhibitor. Dpp-7 inhibitors are useful for  
 CC reducing the growth of a bacterium and protecting an animal from a  
 CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis  
 CC or periodontitis). The present DNA sequence encodes the Porphyromonas  
 CC gingivalis dipeptidylpeptidase-7 (Dpp-7) enzyme of the invention.

XX Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other;

#### Alignment Scores:

Pred. No.: 0 Length: 2139  
 Score: 3719.00 Matches: 712  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 24 Gaps: 0

US-10-008-355-2 (1-712) x AAL43635 (1-2139)

Qy 1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20  
 Db 1 ATGCAATGAATTAATAAGATATTCTCTCGAGCAGCCCTGCTTGGGTGCTTCAGGG 60  
 Qy 21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluLeuAsp 40  
 Db 61 GTAGCAAGGCGCAAGGATGTGGTCTCTCAACGAATCAATCAGGAGAAATCTGGAT 120  
 Qy 41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60  
 Db 121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTGCTCTACAGTTTCGCAAGCGG 180  
 Qy 61 SerIleAlaAsnAlaValIlePheGlyGlyCysThrGlyIleThrValSerAsp 80  
 Db 181 TCCATTGCAATGCGGTGTTATCTTCGGTGGCGGATGACCGGTATCACAGTGTCCGAT 240  
 Qy 81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100  
 Db 241 CAGGCGCTGATCTTTACCAACACCACTACGCGGATACGGTGTATCCAGAGCAAGCAG 300  
 Qy 101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluLeuLeuPro 120  
 Db 301 GTGGATACGACTATCTCGCGGATGGTTTCGTTTCTCGCAGATGGGTGAGGAGCTTCCG 360

Qy 121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140  
 Db 361 ATTCGGGTCTTTCCGTCGAAGTATCTGCGCAAGATCGTGAAGGTAACGCAAGTAGAA 420  
 Qy 141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160  
 Db 421 GCACAGCTCAAGGGTATCACTGACGAGATGGAGGCTCTGCGCAAGCTCAGAGGTATGC 480  
 Qy 161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180  
 Db 481 CAAGACTGGCCCAAAAAGAAAATGCAGACGAGAACCACTCTGCATGCTASAGCCTTTC 540  
 Qy 181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200  
 Db 541 TATTCCAACAACGAATACTTCTCATCTACGATGTATTCAAGGACGTTTCGTATGGTA 600  
 Qy 201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220  
 Db 601 TTTGCTCTCTCCAGCTCTGTAGTAAAGTTTCGGAGCGGATACGACAACTGGATGTGGCGG 660  
 Qy 221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240  
 Db 661 CGTCACAGGGCGACTTCAGCGTATTCCGCGTGTATGCCGTCGCGACACCGCGCGGCC 720  
 Qy 241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGln 260  
 Db 721 GAATACAGCAAGGACAATAAACCTATAAGCCGCTTTACTTCGCTGCGGTATCCATGCAA 780  
 Qy 261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280  
 Db 781 GGCTACAGGCTGACGACTATGCCATGACCATCGCTTTCCCGGCGCAGTACGGATCGCTAC 840  
 Qy 281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300  
 Db 841 CTCACTCTTCGGGTGGGAGATCGTATCGAACAACGAGAACAACTCTCGTATCGAAGTT 900  
 Qy 301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320  
 Db 901 CGCGTATCAAGCAAGGCATCTGGAAGGAAGCCATGAGCGCAGATCAGGCTACCGGTATC 960  
 Qy 321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340  
 Db 961 AAATATGCCAGCAAGTATGCTCAGAGTCTAACTATTTCGAAGAATTCGATCGGTATGAAC 1020  
 Qy 341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArgAlaPheAla 360  
 Db 1021 CGCGTCTCGCTCTGACGTGATAGTCTGTAAGCGTCGCGAGGAAGAGCATTCGCA 1080  
 Qy 361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380  
 Db 1081 GACTGGATCCCTAAGACGCGCAAGAGTCTGTCTATGGCGATGATTGCTCTCTCTCGAA 1140  
 Qy 381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400  
 Db 1141 AAGGCTTATAAGGAAGGAGCAAGCCCAACCGTGAGTGTGACTTATTTCGAGGAGACGCTC 1200  
 Qy 401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420  
 Db 1201 TTCGGTGGTACCGAGGTGGTTCGTTTCACAGTTCGCAACGCAATTCGCTACAAATCCT 1260  
 Qy 421 AspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSer 440  
 Db 1261 GATGCTCATGCGGTATCTCAAAATCGTTGACGCAACAAGTACAAAGACTACTCTCCCTCCG 1320  
 Qy 441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp 460  
 Db 1321 CTCGACCGTAAAGGTGCTCCCGCATGCTCGATATTGTCGCGCGCGGTATCCCTGCCAC 1380  
 Qy 461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480  
 Db 1381 AAGCTCCCGATATATTCAAGAATGTATTCGCAACAAGAATTCAAAGGCGACACGAGAG 1440  
 Qy 481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500



```
Qy 131 LysIleValLysValThrAsp-----LysValGluGlyGln----- 142
::: ||||| ||| ||||| ||||| |||||
Db 2215 CAGACTGTAAAGTTGTTGATGTAATGTGCAAGTAATGGGGCGCATACCCAGAAAT 2274
Qy 143 -----LeuLysGlyIleThrAspGluMet 150
::: ||||| ||||| |||||
Db 2275 ATCGTTATGAAATGGCAGCTGAAATACACACAACCTCAAAAGGGCCAGAAAATTCGCGA 2334
Qy 151 GluArgLeuArgLysAlaGlnGlnValCysGlnGlnLeuAlaLys-----LysGlu 167
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2335 GAGGTTATGAACCAATCGAAGAGGTATCAAGAATTTAGGAAGTCAGATTCCTTAAGAG 2394
Qy 168 AsnAlaAspGluAsn---GlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyr 186
::: ||||| ||||| ||||| ||||| |||||
Db 2395 CATCCTGTTGAAATACTCAATTG-----CAAAATGAAAG 2430
Qy 187 PheLeuIleValTyr-----AspValPheLysAspValArgMetValPhe 201
::: ||| ||||| ||||| ||||| ||||| |||||
Db 2431 CTCATCTCTATTATCTCCAAAATGGAGAGACATGTATGTGGACCAAGAATTAGATATT 2490
Qy 202 AlaProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArg 221
::: ||||| ||||| ||||| ||||| |||||
Db 2490 ----- 2490
Qy 222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlu 241
||| ||||| ||||| ||||| ||||| |||||
Db 2491 -----AATCGTTTAAGTGAT 2505
Qy 242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGly 261
||| ||||| ||||| ||||| ||||| |||||
Db 2506 TATGATCTCGATCAC-----ATTGTTCCCACAAGT 2535
Qy 262 TyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu 281
||| ||||| ||||| ||||| ||||| |||||
Db 2536 TTCCTTAAAGCGAT-----TCAATAGACAATAAGGTC 2568
Qy 282 ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArg 301
::: ||||| ||||| ||||| ||||| |||||
Db 2569 TTAACGGCTTCGATAAAATCGTGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTC 2628
Qy 302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLys 321
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2629 AAAAGATGAAAACACTATTGGAGACAACCTCTAAACGCCCAAGTTAATCACCACGTAAG 2688
Qy 322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341
||| ||||| ||||| ||||| ||||| |||||
Db 2689 TTT-----GATAATTTAAGCAAGCTGAACGTGGA----- 2718
Qy 342 GlyLeuAlaArgLeuAspValIleGly-----ArgLysArgAlaGluGluArgAla 358
||| ||||| ||||| ||||| ||||| |||||
Db 2719 GGTGTTGAGTGAACCTGTATAAAGCTGGTTTATCAACGCCCAATTTGGTTGAAACTCGCAA 2778
Qy 359 PheAlaAspTrpIle-----ArgLysAsnGlyLys----- 368
||| ||||| ||||| ||||| |||||
Db 2779 ATCACTAAGCATGTGGCACAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAT 2838
Qy 368 ----- 368
Db 2839 GATAACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAATTAGTTCTGACTTC 2898
Qy 368 ----- 368
Db 2899 CGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAAACAATTACCATCATGCCCATGAT 2958
Qy 368 ----- 368
Db 2959 CGGTATCTAATGCCGCTGTGGAACTGCTTTTGATTAAAGAAATATCCAAACTTTGAATCG 3018
Qy 369 SerAlaValTyrGly-----AspValLeuSerSerLeuLysAlaTyr 383
||| ||||| ||||| ||||| ||||| |||||
Db 3019 GAGTTTGTCTATGTGATTATAAAGTTTATGATGCTTCGTRAAAGTGAATGCTTAAGCTCGAG 3078
Qy 384 LysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu---PheGly 402
```

```
Db 3079 CAAGAAATAGCAAGCAACGCAAAATATTTCTTTACTCTAATATCATGAATCTCTTC 3138
Qy 403 GlyThrGluVal-----ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAs 421
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3139 AAAACAGAAATTTACACTTGCNAATGGAGAGATTCGCAACGCCCTCTA-ATCGAAACTAA 3197
Qy 421 PAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLe 441
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3198 TGGGAAACTGGA-----GAATTTGCTGGGATAAAGGGCGAGATTTT-----GCCAC 3245
Qy 441 uAspArgLysValPro-----AlaMetLeuAspIleValArgArg----- 456
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3246 AGTCGCAAGATATTGTCCATGCCCCCAAGTCAATATTGTCAGAAAAACAGAGTACAGAC 3305
Qy 457 -----IlePro-----AlaAspLysLeuProAspIle 465
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3306 AGCGGATTTCTCCAGGAGTCAATTTTACCAGAAAAGAAATTCGGACAGCTT----- 3357
Qy 465 ePheLysAsnValIleAspLysPheLysGlyAspThrLysLysTyrAlaAspPheVa 485
||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3358 -----ATTGCTCTGTAATAAAGACTGGGATCCCAAAAAAATATGTTGGT----- 3399
Qy 485 lPheAspLysSerValValProTyrSerAspLysPheIleAla----- 499
||| ||||| ||||| ||||| ||||| |||||
Db 3400 -TTTGATAGTCCAACGGTAGCTTATTCAGTCTTAGTGTGCTAAGTGGAAAAAGGAA 3458
Qy 500 -----MetLeuLysSer-----MetAspLysGluLys 508
||| ||||| ||||| ||||| ||||| |||||
Db 3459 ATCAAGAAAGTTAAATCCGTTAAAGAGTTACTAGGATACAAATATGGAAGAGATTC 3518
Qy 508 sPheAla-----LysAlaIleGluLysAs 516
||| ||||| ||||| ||||| ||||| |||||
Db 3519 CTTTGAANAATCCGATTGACTTTTAGAAGCTTAAGGATATAGGAGTTAAAAAGA 3578
Qy 516 pProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAl 536
||| ||||| ||||| ||||| ||||| |||||
Db 3579 CTTAATCATTAACCTACCTAAA----- 3600
Qy 536 aMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuArgGl 556
||| ||||| ||||| ||||| ||||| |||||
Db 3601 -----TATAGTCTTTTGGATTAGAAAACGGTCGTAACGGGATGCTGCTAGTCGGCGGA 3656
Qy 556 uMetTyrProGlyArg-----AlaLeuProSerAsp---AlaAsnPheThrMetArgMe 573
||| ||||| ||||| ||||| ||||| |||||
Db 3657 ATTACAAAAGGAAATGAGCTGGCTCTGCCAAGCAATATCTGAATTTTATATTTAGC 3716
Qy 573 tSer---TyrGlySerIleLysGlyTyrGluProGlnAspGly-----AlaTr 588
||| ||||| ||||| ||||| ||||| |||||
Db 3717 TAGTCATTATGAAAGTTGAAGGT---AGTCCAGAAGATAACGAACAAAACAAATTTGTT 3773
Qy 588 pTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGl 608
||| ||||| ||||| ||||| ||||| |||||
Db 3774 TGTGGACAGCAT-----AAGCATATTATAGATGAGATTATTGAGCAAAATCAGTGA 3824
Qy 608 uPheAlaValGln----- 612
||| ||||| ||||| ||||| |||||
Db 3825 ATTTTCTAAGCGTCTTATTTAGCAGATGCCAATTTAGATAAAGTCTCTTAGTCGATATAA 3884
Qy 613 -----GluAsnIleLeuAspLeuPheArgTh 621
||| ||||| ||||| ||||| ||||| |||||
Db 3885 CAACATAGACAAACCAATACGTGAACAGCAAGAAAATATTATTCATTTATTATTAGCTT 3944
Qy 621 rLysAsnTyrGly----- 626
||| ||||| ||||| ||||| |||||
Db 3945 GACGAATCTTGGAGCTCCGCTGCTTTTAAATATTTTGTATACAACAATTTGATCGTAAACG 4004
Qy 626 gTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly 645
||| ||||| ||||| ||||| ||||| |||||
Db 4005 ATATACGCTCTACAAAAGAGTTTGTAGATGCCACTCTTATCCATCAATCCATCCTGCT 4062
RESULT 3
AAX91754
```

ID AAX91754 standard; DNA; 2853 BP.  
 XX AC AAX91754;  
 XX DT 25-AUG-1999 (first entry)  
 XX DE Porphyromonas gingivalis protein PG67 encoding DNA.  
 XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
 XX KW vaccine; antigenic; ds.  
 XX OS Porphyromonas gingivalis.  
 XX XX W09929870-AL..  
 XX PD 17-JUN-1999.  
 XX PF 10-DEC-1998; 98WO-AU01023.  
 XX PR 04-AUG-1998; 98AU-0005028.  
 XX PR 10-DEC-1997; 97AU-0000839.  
 XX PR 31-DEC-1997; 97AU-0001182.  
 XX PR 30-JAN-1998; 98AU-0001546.  
 XX PR 10-MAR-1998; 98AU-0002264.  
 XX PR 09-APR-1998; 98AU-0002911.  
 XX PR 23-APR-1998; 98AU-0003128.  
 XX PR 05-MAY-1998; 98AU-0003338.  
 XX PR 22-MAY-1998; 98AU-0003654.  
 XX PR 29-JUL-1998; 98AU-0004917.  
 XX XX (CSLC-) CSL LTD.  
 XX PA Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;  
 XX PI Ross BC, Rothel LJ, Webb EA;  
 XX XX WPI: 1999-385613/32.  
 XX DR P-PSDB: AAX34536.  
 XX PT Antigenic Porphyromonas gingivalis peptides for preventing  
 XX PT gingivitis  
 XX PS Claim 12; Page 244-245; 588pp; English.  
 XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to  
 CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the  
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
 CC activity with a vaccine mechanism of action. The PG polypeptides can be  
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
 CC be used to detect Porphyromonas gingivalis in standard hybridisation  
 CC assays. Porphyromonas gingivalis is involved in periodontal disease  
 CC especially gingivitis.  
 XX SQ Sequence 2853 BP; 736 A; 685 C; 691 G; 741 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.00909 Length: 2853  
 Score: 134.50 Matches: 105  
 Percent Similarity: 34.41% Conservative: 66  
 Best Local Similarity: 21.13% Mismatches: 202  
 Query Match: 3.62% Indels: 124  
 Ds: 20 Gaps: 23  
 US-10-008-355-2 (1-712) x AAX91754 (1-2853)  
 QY 221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArg-ProAl 240  
 Db 717 CTCACAGTCGGTCAATCTGGTATCCCGAATATCGCGTATGAGTCTTCCTTT 776  
 QY 240 aGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGl 260  
 Db 777 CGGCTATTTT-----CCCTTTCCNACAAATACTCTCCGGTATATTGATGCC 824

QY 260 nGlyTyrLysAlaAspPheTyrAlaMetThrIleGlyPheProGlySerThrAspArgTy 280  
 Db 825 CACGTACGGAGAGACAAT-----CGCTATCGATT-----TA 857  
 QY 280 rLeuThrSerTrpGly-----ValGluAspArgIleGluAsnGluAsnPr 296  
 Db 858 TTTGAGGAATGGTGGATATTTTTCGCTTCAGCAGCTATATATCGAT----- 903  
 QY 296 oArgIleGluValArgGly-----IleLysGlnGlyIleTrpLysGluAlaMetSerAl 314  
 Db 904 ----TTGGCATTTGGTGGGAGATCTTTTCCAAAGGTCATGGGGCATTTTCAGCCCAATC 959  
 QY 314 aAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLy 334  
 Db 960 GAAATATAAGAGAGGTATAAGTACACGGCTCTTC-----GAAGCCAATTAT----- 1008  
 QY 334 sAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAl 354  
 Db 1008 ----- 1008  
 QY 354 aGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAs 374  
 Db 1009 -----CTGGTATCGAAGTCCGGCGACAAATACGTGCCCGGAGA 1046  
 QY 374 pValLeuSer-----SerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAs 390  
 Db 1047 CTACAGCAAGACCACCATCTGTAATATCCGATGGACACACAGTCAGGATCCGAAGGCCAA 1106  
 QY 390 nArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAl 410  
 Db 1107 TCCT-----TTGCAACAGTTGTCCGCCAATGTCAATTTTGCACCGGGAG 1151  
 QY 410 aGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla-----GlyIleLeuLy 428  
 Db 1152 CTATTTCCAGATTCGCTGATATACCATCTATGATGTCAATCCCTACTGTCTACGACACG 1211  
 QY 428 sSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAl 448  
 Db 1212 AAGTTCGGCCGTGAGCTATTCGCGCAAGTTTCCGGGTACTCCTTTTTCGATTACGGGTAG 1271  
 QY 448 aMetLeuAspIle-----ValArgArgArgIleProAlaAspLysLeuProAspIl 465  
 Db 1272 CATG---GATATCAGCCAGAACATCGCGGATACGAGGTGAGCCTTACTTTGCGCAATCT 1328  
 QY 465 ePheLysAsnValIleAspLys---LysPheLysGlyAspThrLysLysTyrAlaAspPh 484  
 Db 1329 TTCGATTAATATGTCGCGCGTATCTTTCACGGGAGACCGGTGTAGACCGGACG 1388  
 QY 484 eValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMe 504  
 Db 1389 ATGTTACGAGAAGTTGAGTGTGGCTATTTCGGTCAGCTTCGCAATAGTATCTTTGACAAA 1448  
 QY 504 tAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSe 524  
 Db 1449 AGAAGAAGATTTGCTCCAGAGCAATCTCTGCGGATTTGGAGAATGTTATGCGTCATTC 1508  
 QY 524 rValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGl 544  
 Db 1509 CGTACCGATCAGTTGATGCTCCCTTTTCTGGATTATATCAATCTGATGATGGGGTTAA 1568  
 QY 544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyr-----ProGlyAr 561  
 Db 1569 CTACAATGAGTGGTGTACACAGAAAGCATACGGAAGTCGCGAATGAGGATAGAGAAAC 1628  
 QY 561 gAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTy 581  
 Db 1629 ATTCTGCTTCCGGACAGCACTATAAATTCGCGAGACTGTAC-----GATTA 1676  
 QY 581 rGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLy 601  
 Db 1677 CAGTCTGTGCGCAGGC-----TTATCTACCACATTTGACGATGATCTTCAAGCC 1724  
 QY 601 sGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu----- 618

```

Db 1725 TTGGAACCT-----TTTTCTCGGAGCAATCTCATATGATCGCGTCATCG 1772
Qy 619 -----PheArgThrLysAsnTyrGlyAr 626
Db 1773 CTTACGCCCACTGTCAGTTCTTCCTATATGCGGACTTCACGAACCGCATATGGCTT 1832
Qy 626 gTyrAla-----GluAsnGlyGlnLeuHisIleAlaPheLeuSerAs 640
Db 1833 TTGGGAGCTTCTTGAGCATACGGATCAGACCGCAAGCTGCATACGCTGCTACTCTCC 1892
Qy 640 n-----AsnAspIleThrGly-----GlyAsnSerGlySerProValPheAs 654
Db 1893 TTATTTTCGAGCAGATATTCGGTCTCCCTCCATGCGCAATGCAGGATCT----- 1941
Qy 654 pLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAla 670
Db 1942 -----GTCAATTTCTCTTTTGACAACTTAGAGGCC 1974
RESULT 4
AAX91621
ID AAX91621 standard; DNA; 2859 BP.
XX AC AAX91621;
XX DT 25-AUG-1999 (first entry)
XX DE Porphyromonas gingivalis protein PG67 ORF encoding DNA.
XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX KW vaccine; antigenic; ds.
XX OS Porphyromonas gingivalis.
XX PN WO929870-A1.
XX PD 17-JUN-1999.
XX PF 10-DEC-1998; 98WO-AU01023.
XX PR 04-AUG-1998; 98AU-0005028.
XX PR 10-DEC-1997; 97AU-0000839.
XX PR 31-DEC-1997; 97AU-0001182.
XX PR 30-JAN-1998; 98AU-0001546.
XX PR 10-MAR-1998; 98AU-0002264.
XX PR 09-APR-1998; 98AU-0002911.
XX PR 23-APR-1998; 98AU-0003128.
XX PR 05-MAY-1998; 98AU-0003338.
XX PR 22-MAY-1998; 98AU-0003654.
XX PR 29-JUL-1998; 98AU-0004917.
XX PA (CSLC-) CSL LTD.
XX PI Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
XX PI Ross BC, Rochel LJ, Webb EA;
XX WPI; 1999-385613/32.
XX DR P-PSDB; AAY34403.
XX PT Antigenic Porphyromonas gingivalis peptides for preventing
XX PT gingivitis
XX PS Claim 12; Page 153-154; 588pp; English.
XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX CC activity with a vaccine mechanism of action. The PG polypeptides can be
XX CC used as vaccines especially against Porphyromonas gingivalis. Probes can
XX CC be used to detect Porphyromonas gingivalis in standard hybridisation
XX CC assays. Porphyromonas gingivalis is involved in periodontal disease
XX CC especially gingivitis.

```

```

XX SQ Sequence 2859 BP; 738 A; 687 C; 692 G; 742 T; 0 other;
Alignment Scores:
Pred. No.: 0.00912 Length: 2859
Score: 134.50 Matches: 105
Percent Similarity: 34.41% Conservative: 66
Best Local Similarity: 21.13% Mismatches: 202
Query Match: 3.62% Indels: 124
Db: 20 Gaps: 23
US-10-008-355-2 (1-712) x AAX91621 (1-2859)
Qy 221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArg-ProAl 240
Db 723 CGTCACAGTCCGGTCAATCTGGTTATCGCGCATATGCGGTCGCGATAGTCTTCCTTT 782
Qy 240 aGlyTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaLavalSerMetGl 260
Db 783 CGGCTATTTT-----CCCTTTTCCAACAAATACTCTTCCGGTATATTGATGCC 830
Qy 260 nGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTy 280
Db 831 CACGTACGAGAGGACAAT-----CGCTATGGAATT-----TA 863
Qy 280 rLeuThrSerTrpGly-----ValGluAspArgIleGluAsnGluAsnAsnPr 296
Db 864 TTGAGGAATGCTGATATATTGCTTCAGGACATATATCGAT----- 909
Qy 296 oArgIleGluValArgGly-----IleLysGlnGlyIleTrpLysGluAlaMetSerAl 314
Db 910 ----TTGGCATTTGCGTGGGAGAGTCTTTTCCAAGGGTCATGGGCGATTTCAGCCCAATC 965
Qy 314 aAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLy 334
Db 966 GAAATATAAGAAGAGGTATAAGTACACAGCGCTCGTTC-----GAAGCCAAATAT- 1014
Qy 334 sAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAl 354
Db 1014 ----- 1014
Qy 354 aGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAs 374
Db 1015 -----CTGGTATCGAAGTCCGGCGACAAATACGTGCCCGGAGA 1052
Qy 374 pValLeuSer-----SerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAs 390
Db 1053 CTACAGCAAGACCACCACTGCTGAATATCCGATGGACACACAGTCAGGATCCGAAGGCCAA 1112
Qy 390 nArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAl 410
Db 1113 TCCT-----TTGCAACGTTGTCGCCCAATGTCATTTTGCACCGCGAG 1157
Qy 410 aGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla-----GlyIleLeuLy 428
Db 1158 CTATTTCCAGAAATTCGCTGAATACCACTATGATGTCAATGCCGCTACTGTACGACACG 1217
Qy 428 sSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAl 448
Db 1218 AAGTTCGGCGCTGAGCTATTCGCGCAAGTTTCCGGGTACTCCTTTTTCGATTACGGGTAG 1277
Qy 448 aMetLeuAspIle-----ValArgArgArgIleProAlaAspLysLeuProAspIle 465
Db 1278 CATG---GATATCAGCCAGACATGCGCGATACGCGGTGAGCCTTACCTTGCCGAATCT 1334
Qy 465 ePheLysAsnValIleAspLys----LysPheLysGlyAspThrLysLysTyrAlaAspPh 484
Db 1335 TTCGATTAATATGTCCACGCGCTTATCTCTTCAACGGGAAGACCGCTGTAGGACCGCGC 1394
Qy 484 eValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMe 504
Db 1395 ATGGTACGAGAAGTTGAGTGTGGGCTATTTCGGGTTCGCAATAGTATCTTTCACAAA 1454

```



QY 504 tAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSe 524  
Db 1455 AGAAGAAGATTGCTCCAGAGCAATCTCGTGGCGATTGGAAGAAATGGTATCGTCATTC 1514  
QY 524 rValIleAlaAlaAraArgAlaIleGluAlaAspAlaMetAlaAsnAlaTyrAlaIleGlu 544  
Db 1515 CGTACCGATCAGTTGACGTGCTCCCTTTGTTGATTATATCAATCTGACATCGGGGTAA 1574  
QY 544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyr-----ProGlyAr 561  
Db 1575 CTACAATGAGTGGTGTACACCAAGGACATACCGAAGTCGTGGAATGAGGATAAGAAAC 1634  
QY 561 gAlaLeuProSerAspAlaAsnPhetHrMetArgMetSerTyrGlySerIleLysGlyTy 581  
Db 1635 ATTCTGCGCTTCGGACAGCATTATAAATTCGCGACACTGTAC-----GATTA 1682  
QY 581 rGluProGlnAspGlyAlaTyrTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLys 601  
Db 1683 CAGTCTGTCGGCAGGC-----TTATCTACCACATCTACGGTATGTTCAAGCC 1730  
QY 601 sGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu----- 618  
Db 1731 TTGGAACCT-----TTTTCCTCGGAGCAATCTCATATGATCCGTCATCG 1778  
QY 619 -----PheArgThrLysAsnTyrGlyAr 626  
Db 1779 CTTACGCCCACTGTCAGTTCTCTCTATATGCGGACTTCACCAACGCGCATATGGCTT 1838  
QY 626 gTyrAla-----GluAsnGlyGlnLeuHisIleAlaPheLeuSerAs 640  
Db 1839 TTGGGAGCTCTTGACATACGATACGATCAGAACGCGACGCTGTCATACGCTCTCTACTCTCC 1898  
QY 640 n-----AsnAspIleThrGly-----GlyAsnSerGlySerProValPheAs 654  
Db 1899 TTATTTTCGAGCAGATATTCGGTGTCTCCCTCCATGGCAATGCGAGGATCT----- 1947  
QY 654 pLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAla 670  
Db 1948 -----GTCAATTTCTCTTTTGACACAACTTAGAGGCC 1980  
RESULT 5  
AAS55647  
ID AAS55647 standard; DNA; 2652 BP.  
XX AC AAS55647;  
XX AC  
DT 13-FEB-2002 (first entry)  
XX DE Streptococcus pneumoniae DNA for cellular proliferation protein #218.  
XX KW Antisense; ds; prokaryotic cellular proliferation gene;  
KW KW antibiotic; antibacterial; drug design.  
XX OS Streptococcus pneumoniae.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US09180.  
XX PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-267931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.  
DR P-PSDB; AAU37788.  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids  
XX  
PS Claim 27; Seq ID No 9284; 51pp; English.  
XX  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
CC pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence encodes an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2652 BP; 736 A; 609 C; 618 G; 589 T; 0 other;

## Alignment Scores:

Pred. No.: 0.0115 Length: 2652  
Score: 133.00 Matches: 106  
Percent Similarity: 33.20% Conservative: 64  
Best Local Similarity: 20.70% Mismatches: 164  
Query Match: 3.58% Indels: 178  
DB: 23 Gaps: 25

US-10-008-355-2 (1-712) x AAS55647 (1-2652)

QY 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203  
Db 76 GATGTTTTCAGCCCTCAGCGCATCAAAAGCTATGACCTTATTCATCGCTTATTCACCA 135  
QY 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223  
Db 136 CCAACGTTACAGGTAACCTT-----CACCTT 162  
QY 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243  
Db 163 GGT-----CACGCT 171  
QY 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263  
Db 172 TGGATACA---ACTTTGCAAGATATTATCATCCGTCFAAAACGCATGCAAGGTTTTGAT 228  
QY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282  
Db 229 -----ACCTTTGGCTTCTGGGATGGACACGACGGGATGCGCACT 270  
QY 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302  
Db 271 CAGGCTAAGTTGAGGAGCGCTTGGCTGGTGGAGGCAATTTCCCGCATGACCTTGGTCTGT 330  
QY 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318  
Db 331 GAGTCTTTCTTGACGAAGTCTGGGAATGGAAGAC-----GAATATGCCACT 378  
QY 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337  
Db 379 ACCATCAAGGAACAATGGGGCAAGATGGGGCTCTCTGTAGACTATTCTCTGATGACGCTTTC 438

```
QY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArg 357
Db 439 ACTCTTGACGAGGTTGTCA-----AAAGCTTTCGTAAAG 474
QY 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
Db 475 GTCTTTGTGGACCTTTACAAAGAGCGTGATCTACCGTGGTGAGTTATCATCAACTGG 534
QY 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
Db 535 GACCCAGCAGCTCGCACAGCCCTTTCTGATATTGAGGTGATTCAACAGGATGTGAAGGT 594
QY 385 -----GluGlyAlaLysAlaAsnArgGluMetThr 394
Db 595 GCCTTCTACCATGATTTACATGCTGGAGATGTTTCACGGCGCCTTCAAGTTGCTACA 654
QY 395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
Db 655 ACTCGTCTGAGACTATGTTGGGACGTT----- 684
QY 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434
Db 685 GCGATTGCGGTTAATCCAGAA-----GACCCGCGCTAC 717
QY 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 718 AGGACTTGATTGGTAAACACGTCCTCTCCA---ATCGCTAATAAAGCTCATCCCAATC 774
QY 449 MetLeuAsp-----IleValArgArg 455
Db 775 GTTGAGATGAGCAGCAGCATCCTGAGTTGTTGCTACTGTTGCTGTAATAATCACACTGCC 834
QY 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
Db 835 CAGGATCCAAATGACTCTTGTGTCGCCAACGTCATATCTGCCCAAGCAAGCAATCATG 894
QY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db 895 AAGCAGCAGGAACTATGAACGAGCTGTGCTTTGAATTTTCAGGCATG-----GAT 945
QY 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
Db 946 CGTTTTTGAAGCTCGT-----AAGGCAGCTGTTGCTAAGTTGGAAGAAATC 990
QY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
Db 991 GGTGCCCTTGTCAAAATCAAAACGTGCCACAGTGTGGTCACTCAACAGCTACAGGT 1050
QY 532 -----IleGlnAla 534
Db 1051 GTCGTAGTTGAGCAGCGCTTGTCTACTCAATGTTGCTCAAGATGGACCAATTGGCTAAG 1110
QY 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeu 554
Db 1111 AACGCCATTGCCAACCAACAGACACAGAGGACAAG----- 1143
QY 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565
Db 1144 GTCGAATTTCTACCCACCTCGTTTCAACGATACCTTCTCAATGATGATGAAATGTCCAC 1203
QY 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
Db 1204 GACTGGTTTATCTCTCGTCAGCTCTGTTGGGGGTAC-----CAAAATC 1245
QY 586 GlyAlaTrpTyrAsnTyrHisThrGlyLysGlyValLeuLysGlnAspProLys 605
Db 1246 CTGCTGTGTACAAT-----GCTGATGTTGAATGTATGTGCGGGAAGAACTCCAGAA 1299
QY 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
Db 1300 GTGACGGATGACTCAGGACCAAGACGCTCTGGAT 1335
RESULT 6
```

```
AAAX57027
ID AAX57027 standard; DNA; 2652 BP.
AC AAX57027;
XX
XX 19-JUL-1999 (first entry)
XX
XX S. pneumoniae vals coding region DNA.
XX
XX vals; treatment; antagonist; ss.
XX
XX Streptococcus pneumoniae.
XX
XX JP11113577-A.
XX
XX 27-APR-1999.
XX
XX 17-OCT-1997; 97JP-0321886.
XX
XX 17-OCT-1997; 97JP-0321886.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX
XX WPI; 1999-320830/27.
XX
XX P-PSDB; AAY08339.
XX
XX New vals polypeptide and polynucleotide - useful for treatment of
XX conditions associated with abnormal vals protein levels
XX
XX Claim 4; Page 24; 27pp; Japanese.
XX
XX This invention describes the isolation of a novel Streptococcus
XX pneumoniae vals protein. The new polypeptide and antagonist are
XX useful for the treatment of an individual with abnormal vals
XX protein levels, by administering vals to individuals requiring
XX vals polypeptide, and administering the antagonist to individuals
XX requiring inhibition of vals polypeptide.
XX
XX Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.0223 Length: 2652
Score: 130.00 Matches: 106
Percent Similarity: 32.81 Conservative: 62
Best Local Similarity: 20.70 Mismatches: 166
Query Match: 3.50 Indels: 178
Db: 25
US-10-008-355-2 (1-712) x AAX57027 (1-2652)
QY 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
Db 76 GATGTTTCAACCTTCAGCGCATCAAAGCGTAAAGCTTATCAATCGTTATTCACCA 135
QY 204 ProSerSerValGlyLysPheGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
Db 136 CCAACGTTACAGGTAACCTT-----CACCTT 162
QY 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
Db 163 GGT-----CACGCT 171
QY 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
Db 172 TGGGATACA---ACTTTCAGATATTTATCATCCCTCAAAACGATCAAGGTTTGTAT 228
QY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
Db 229 -----ACCCCTTTGGCTTCTCGGTGGATGGACACGAGGATTCGCACT 270
QY 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302
Db 271 CAGGCTAAGGTAGAGGAGCGCTTGGCTGTGGGCAATTTCCCGCTATGACCTTGTGCT 330
```

```
Qy 303 -----IleLysGlnGlyIleTyrLysGluAlaMetSerAlaAspGlnAlaThr 318
Db 331 GAGTCTTCTTGACGAAGCTGGGAATGGAAGAC-----GAATATGCCACT 378
Qy 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTyrLysAsnSerIle 337
Db 379 ACTATCAAGGAAACAAATGGGCAAGCTCTCTAGACTATTCTCGTAGCGGTTTC 438
Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
Db 439 ACTCTTGACGAAGCTTGTCA-----AAGCTGTCGTAAG 474
Qy 358 AlaPheAlaAspTyrPheArgLys-----AsnGlyLysSerAlaValTyrGly 373
Db 475 GTCCTTGTGACCTTTACAGAAAGGCTGATCTACCGTGTGAGTTTATCATCACTG 534
Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
Db 535 GACCCAGAGCTCCACAGCCCTTTCTGATATTAGGTGATTCACAGGATGTGGAAGGT 594
Qy 389 AlaAsnArgGluMetThrTyrLeu----- 396
Db 595 GCCTTCTACCATGATGATTACATGCTGGAAGATGTTACCGCTCCTGAAGTTGCTACA 654
Qy 397 -----SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
Db 655 ACTCGTCTGACACCATGTTTGGGGACGTT----- 684
Qy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
Db 685 GCGTGTGCGGTCAACCCAGAA-----GACCCGCGGTAC 717
Qy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 718 AAGACTTGATTGTTAAATAATGTCATCTTCCA---ATCGCTAATAAATCATCCCAATC 774
Qy 449 MetLeuAsp-----IleValArgArg 455
Db 775 GTTGAGATGAGCAGCAGATCCTGAGTTGTGCTGCTGCGTGAATAATCACACCTGCC 834
Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
Db 835 CACGATCCAAATGACTCTTGGTTGGCCAGCTGATCAATCTGCCACAGTCAACGTCATG 894
Qy 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValProTyrSerAsp 495
Db 895 AACGACGACGGAACCATGATGACTTGCCTTTGAAATTTTCAGGCATG-----GAC 945
Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
Db 946 CGTTTGAAGCTCGT-----AAGGCAGTCGTTGCTAAGTTGGAAGAAATC 990
Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
Db 991 GGTGCCCTCGTCAAAATGGAATAACGTTCCACAGTGTGCTCACTCAGACGCTACAGGT 1050
Qy 532 -----IleGlnAla 534
Db 1051 GTTGTGGTTGAACCTCGCTTCTACTCAATGCTGCTCAAGATGGACCAATGGCTTAAG 1110
Qy 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
Db 1111 AACGCCATTGCAACCAAGACACAGAGGACAAAG----- 1143
Qy 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565
Db 1144 GTCGAAATTCATCCCACTCTGTTTCAACGATACCTCTTCATGATGATGAAATGTCAC 1203
Qy 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGluAsp 585
Db 1204 GACTGGGTTATCTCTGCTCAGCTCTGGTGGGTAC-----CAATC 1245
```

```
Qy 586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
Db 1246 CCTCGCTGGTACAAT-----GCTGATGGTGAATATGATCGGCGAAGAAGCTCCAGAA 1299
Qy 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
Db 1300 GGTACCGATGGACTCAGGACGAAGACGCTCTGGAT 1335
RESULT 7
AAAX32814
ID AAAX32814 standard; DNA: 2652 BP.
XX AAAX32814;
XX 25-JUN-1999 (first entry)
DE S. pneumoniae valyl tRNA synthetase (Vals) polypeptide encoding DNA.
XX Valyl tRNA synthetase; vals; otitis media; conjunctivitis; pneumonia;
KW bacteremia; sinusitis; pleural empyema; endocarditis; meningitis;
KW bacterial adhesion; bacterial infection; ss.
XX Streptococcus pneumoniae.
OS EP909818-A1.
PN 21-APR-1999.
XX 17-OCT-1997; 97EP-0308257.
XX 17-OCT-1997; 97EP-0308257.
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX Brown JR, Jaworski DD, Lawlor EJ, Wang M;
XX WPI; 1999-231674/20.
XX P-PSDB: AAY06828.
XX New Streptococcus pneumoniae valyl tRNA synthetase (vals) gene and
XX protein, useful for screening for antibacterial compounds, for
XX prevention and treatment of Streptococci infections: e.g. meningitis
XX and conjunctivitis
XX Claim 4; Page 6-9; 35pp; English.
XX This DNA encodes a S. pneumoniae vals polypeptide (EC 6.1.1.9), a member
XX of the valyl tRNA synthetase family. The polypeptide can be expressed
XX recombinantly by transforming host cells with vectors comprising the vals
XX gene. Vals polypeptides and polynucleotides are useful for diagnosing
XX diseases due to an infection of an organism and also the stage and type
XX of the infection. The polypeptides are also used for screening for their
XX modulators which can be used in the treatment of diseases. Diseases
XX diagnosed, prevented or treated include otitis media, conjunctivitis,
XX pneumonia, bacteremia, sinusitis, pleural empyema, endocarditis and
XX especially meningitis. Vals polypeptides, polynucleotides and their
XX (antagonists) can prevent adhesion of bacteria to matrix proteins, and
XX are useful for use on wounds and body implants to prevent bacterial
XX infection.
XX Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;
SQ Alignment Scores:
Pred. No.: 0.0223 Length: 2652
Score: 130.00 Matches: 106
Percent Similarity: 32.81% Conservative: 62
Best Local Similarity: 20.70% Mismatches: 166
Query Match: 3.50% Indels: 173
DB: 20 Gaps: 25
US-10-008-355-2 (1-712) x AAAX32814 (1-2652)
Qy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
```

```

Db 76 GATGTTTCAAGCCTTCAGGCGATCAAAAGGCTAAGCCTTATTCAATCGTTATTCCACCA 135
    |||||
Qy 204 ProSerSerValGlyLysPheGlyClyAspThrAspAsnTrpMetTrpProArgHisThr 223
    |||||
Db 136 CCAACGTTACAGGTAAACTT-----CACCTT 162
Qy 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
    |||
Db 163 GGT-----CACGCT 171
Qy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
    |||
Db 172 TGGGATACA---ACTTTCGAAGATATTATCATCCGTCAAAACGATCGCAAGGTTTTCAT 228
Qy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
    |||||
Db 229 -----ACCTTTGGCTTCTCGGATGGACCGACGAGGATGGCCACT 270
Qy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302
    |||||
Db 271 CAGGCTAAGGTAGAGGCGCTTGGGTGCTGAGGCATTTCCCGCTATGACCTTGGTCGT 330
Qy 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
    |||
Db 331 GAGTCTTTCTTGACGAAGTCTGGGAATGGAAGAC-----GAATATGCCACT 378
Qy 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
    |||||
Db 379 ACTATCAAGGAACAAATGGGGCAAGTGGGCTCTCTGTAGACTATTCTCTGGAGCGTTTC 438
Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluLysArg 357
    |||||
Db 439 ACTCTTGACGAAGTTTGCTCA-----AAAGCTGTCGTAAG 474
Qy 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
    |||
Db 475 GTCCTTTGGACCTTTACAAGAAAGGTGATCTACCGTGGTGTGAGTTATCATCAACTGG 534
Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388
    |||
Db 535 CACCCAGCAGCTCGCACAGCCCTTCTGATATTGAGGTGATTCACAGGATGTGGAAGGT 594
Qy 389 AlaAsnArgGluMetThrTyrLeu----- 396
    |||
Db 595 GCCTTCTACCATGAATTACATGCTGGAAGATGTTTCACGCGCTCTTGAAGTTGCTACA 654
Qy 397 -----SerGluThrLeuPheGlyClyThrGluValValArgPheAlaGlnPheAlaAsn 414
    |||||
Db 655 ACTCGTCTCTGAGACCATGTTTGGGACGCT----- 684
Qy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
    |||||
Db 685 CGGTTGCGGTCAACCCAGAA-----GACCCGCGCTAC 717
Qy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
    |||||
Db 718 AAGGACTTGATTGTTGATAAAATGTCATCTTCCA---ATCGCTAATAAACTCATCCCAATC 774
Qy 449 MetLeuAsp-----IleValArgArg 455
    |||
Db 775 GTTGGAGATGAGCAGCAGATCCTGAGTTGGTACTGCTGCTGGAATAATCACACCTGCC 834
Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
    |||
Db 835 CACGATCCAAATGACTCTTGGTGGCCACGTCATTAACCTTGCCCAAGTCAACGTCATG 894
Qy 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
    |||
Db 895 AACGACGAGCGGAACCATGAATGACTTGGCCCTTCAATTTTCAGGCATG-----GAC 945
Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
    |||||

```

```

Db 946 CGTTTGAAGCTCGT-----AAGCAGTCGTTGCTAAGTTGAAAGAAATC 990
Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
    |||||
Db 991 GGTGCCCTCGTCAAAATCGAAAAACGCTGCCACAGTGTGCTACTCAGACGCTACAGGT 1050
Qy 532 -----IleGlnAla 534
    |||
Db 1051 GTTGTGGTTGAACCTCGCTTGTCTACTCAATGGTTCGTCAGATGGACCAATTGGCTAAG 1110
Qy 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
    |||||
Db 1111 AACGCCATTGCCAACCAAGACACAGAGACAAG----- 1143
Qy 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565
    |||
Db 1144 GTCGAATTCACCCACCTCGTTCACAGATACCTTCCTCAATGGATGGAAAAATGTCCAC 1203
Qy 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
    |||
Db 1204 GACTGGGTATCTCTCGTCAGCTCTGGTGGGTCTAC-----CAAATC 1245
Qy 586 GlyAlaTrpTyrAsnTyrHisThrGlyLysGlyValLeuGluLysGlnAspProLys 605
    |||||
Db 1246 CCTGCCTGGTACAAT-----GCTGATGGTGAATGTATGTGCGGGAAGAGCTCCAGAA 1299
Qy 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
    |||
Db 1300 GGTGACGGATGGACTCAGGACGAGACGCTCTTGGAT 1335
    |||||
RESULT 8
AAAL4369
ID AAAL4369 standard; DNA; 2652 BP.
XX
AC AAAL4369;
XX
DT 15-AUG-2000 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae valyl tRNA synthetase (vals).
XX
KW Valyl tRNA synthetase; vals; inhibitor; stringent response;
XX drug screening; antibacterial; antibiotic; genetic immunisation;
XX antibody; bacterial infection; meningitis; ds.
XX
OS Streptococcus pneumoniae strain 0100993.
XX
PH Key Location/Qualifiers
FT CDS 1..1652
FT /*tag= a
FT /product= "Streptococcus pneumoniae valyl tRNA synthetase
FT (vals)"
XX
PN US6051413-A.
XX
PD 18-APR-2000.
XX
PF 24-SEP-1998; 98US-0159539.
XX
PR 18-APR-1996; 96GB-0007791.
XX 17-OCT-1997; 97US-0953492.
XX 18-APR-1997; 97US-0844064.
XX
PA (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI Brown JR, Lawlor EJ, Wang M, Jaworski DD;
XX
DR WPI: 2000-338311/29.
XX P-PSDB; AAY90514.
XX
PT Novel vals polypeptides of valyl tRNA synthetase family useful for
XX treating otitis media, conjunctivitis, pneumonia and bacteremia
XX comprises a specified amino acid sequence -
XX

```

PS Claim 13; Columns 7-10; 2lpp; English.

XX This sequence represents DNA encoding Streptococcus pneumoniae valyl tRNA synthetase (vals, NCIMB Deposit No. 40794). Streptococci are the cause of several types of human diseases, including otitis media, conjunctivitis, pneumonia, bacteraemia, sinusitis, pleural empyema, endocarditis and especially meningitis. The frequency of S. pneumoniae infections has risen dramatically over the past 20 years, probably due to the emergence of multiply antibiotic resistant strains and an increasing population of immunocompromised people. Vals represents a target for new antibacterial agents. Inhibition of tRNA synthetases such as vals leads to a reduction in the levels of charged tRNA, which triggers a cascade of responses (known as the stringent response) resulting in a state of dormancy in the bacterium. Vals, its variants and fragments, anti-vals antibodies, vals inhibitors and nucleotides encoding vals may be used in the diagnosis, prevention and treatment of bacterial infections such as meningitis. Vals can be used to screen compounds for inhibitory activity. Vals may also be useful as an antigen for vaccination of a host to produce specific antibodies which protect against bacterial invasion into damaged tissues. Such antibodies could, for example, prevent the adherence of bacteria to wounds. Nucleotides encoding vals may be used as diagnostic reagents and therapeutic or prophylactic agents, particularly for genetic immunisation.

XX SQ Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;

#### Alignment Scores:

Pred. No.:	0.0223	Length:	2652
Score:	130.00	Matches:	106
Percent Similarity:	32.81%	Conservative:	62
Best Local Similarity:	20.70%	Mismatches:	166
Query Match:	3.50%	Indels:	178
DB:	21	Gaps:	25

US-10-008-355-2 (1-712) x AAA14369 (1-2652)

QY	192	AspValphe-----	-----LysAspValArgMetValPheAlaPro	203
DB	76	GATGTTTTCAGGCTTCAGGCGATCAAAAGCTTAAGCCTTATTCAATCGTTATTCACCA	135	
QY	204	ProSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr	223	
DB	136	CCAAACGTTACAGGTAACCTT-----	CACCTT	162
QY	224	GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer	243	
DB	163	GGT-----	CACGCT	171
QY	244	LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys	263	
DB	172	TGGGATACA---ACTTTGCAAGATATTATCATCGTCAAAACGCATGCAAGGTTTGTAT	228	
QY	264	AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr	282	
DB	229	-----ACCCTTGGCTTCCCTGGGATGACACGACGAGGATGGCCACT	270	
QY	283	SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly	302	
DB	271	CAGGCTAAGGTAGAGAGCGCTTCGCTGGTGAGGGCATTTCCCGCTATGACCTTGGTCGT	330	
QY	303	-----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr	318	
DB	331	GAGTCCTTTCTTGACGAAAGTCTGGGAATGGAAGAC-----GAATATGCCACT	378	
QY	319	ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle	337	
DB	379	ACTATCAGGNACAACTGGGCAAGATGGGCTCTCTAGACTATTCTCTGGACCGTTTC	438	
QY	338	GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg	357	
DB	439	ACTCTTGACGAAGCTTTGTCA-----	AAAGCTGTTTCGTAAG	474
QY	358	AlaPheAlaAspTrpIleArgLys-----	AsnGlyLysSerAlaValTyrGly	373

DB	475	GTCTTTGTGACCTTTTAAAGAAAGGCTGGATCTACCTGCTGGTTTATCATCAACTGG	534	
QY	374	Asp-----ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys	388	
DB	535	GACCCAGCAGCTCCACAGCCCTTCTGATATGAGGTGATTCACAGGATGTGGAAGGT	594	
QY	389	AlaAsnArgGluMetThrTyrLeu-----	396	
DB	595	CCCTTCTACCACTGAATACATCTGGAAGTGTTCACGCGCTCTGAAGTTGCTACA	654	
QY	397	-----SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn	414	
DB	655	ACTGCTCCTGAGACCATGTTTGGGAGCGTT-----	684	
QY	415	AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr	434	
DB	685	CGGTTCCGCTCAACCCAGAA-----	GACCCCGCTAC	717
QY	435	LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla	448	
DB	718	AAGGACTTGATTGGTAAATAATGTCTCTTCCA--ATCGCTAATAAATCATCCCAATC	774	
QY	449	MetLeuAsp-----	IleValArgArg	455
DB	775	GTTGGAGATGAGCAGCAGATCTCTGAGTTTGGTACTGCTGCTGGAATAATCACACCTGCC	834	
QY	456	ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys	475	
DB	835	CAGCATCCAATGACTTCTTGGTGGCCACAGTCATCACTTCCGACAGATCAACGCTCATG	894	
QY	476	GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp	495	
DB	895	AACGACGACGGAACCATGAATGACTTGGCCTTTGAATTTTCAGGCATG-----	GAC	945
QY	496	LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys	515	
DB	946	CGTTTTCAGAGCTCGT-----AAGCAGCTCGTGTCTAAGTTGGAAGAAATC	990	
QY	516	AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla-----	531	
DB	991	GGTGCCCTCGTCAAAATCGAAAACGTCGTCACAGTGTGGTCACTCAGACCGTACAGGT	1050	
QY	532	-----	IleGlnAla	534
DB	1051	GTTGTGTTGAACCTCGCTTCTACTCAATGTTCTGTCGAAGATGGACCAATTGGCTAAG	1110	
QY	535	AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu	554	
DB	1111	AACGCCATTGCCAACCAAGACACAGAGGACAAG-----	1143	
QY	555	ArgGluMetTyrProGlyArgAlaLeuProSer-----	565	
DB	1144	GTCAATCTACCCACCTCGTTTCAACGATACCTTCTCAATGGATGGAAATGTCCAC	1203	
QY	566	AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp	585	
DB	1204	GACTGGCTTATCTCTCGTCAGCTCTGGTGGGTGTCAC-----CAAATC	1245	
QY	586	GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys	605	
DB	1246	CCTGCCCTGGTACAAAT-----GCTGATGTGTAATGTATGTCGGCGAAGAGCTCCAGAA	1299	
QY	606	SerAspGluPheAlaValGlnGluAsnIleLeuAsp	617	
DB	1300	GGTGACCGATGGACTCAGGACGAGAGAGCTCTTGAT	1335	
DB	1300	GGTGACCGATGGACTCAGGACGAGAGAGCTCTTGAT	1335	

RESULT 9  
ABK52414  
ID ABK52414 standard; DNA: 2652 BP.  
XX  
AC ABK52414;  
XX

DT	14-AUG-2002 (first entry)	
XX		
DE	DNA encoding Streptococcus pneumoniae Vals.	
XX		
KW	Vals; microbial infection; antibacterial; gene; ds.	
XX		
OS	Streptococcus pneumoniae.	
XX		
FH	Location/Qualifiers	
FT	CDS 1..2651	
FT	/*tag= a	
FT	/product= "Vals"	
XX		
PN	JP2002119292-A.	
XX		
PD	23-APR-2002.	
XX		
PF	17-OCT-1997; 2001JP-0216648.	
XX		
PR	17-OCT-1997; 97JP-0321886.	
XX		
PA	(SMIK ) SMITHKLINE BEECHAM CORP.	
XX		
DR	WPI: 2002-475332/51.	
DR	P-PSDB; AAU97883.	
XX		
PT	Novel vals polynucleotide useful for producing a vals polypeptide, used	
PT	in the treatment of vals-related disorders	
XX		
PS	Claim 4; Fig 1-3; 24pp; Japanese.	
XX		
CC	The invention describes an isolated vals polynucleotide encoding a mature	
CC	polypeptide the same as that expressed by a vals gene contained in a	
CC	deposited strain of Streptococcus pneumoniae, a complementary sequence or	
CC	a polynucleotide containing at least 15 continuous bases of the	
CC	polynucleotide, the antibody to the encoded protein is used for the	
CC	treatment of microbial infection. This sequence encodes Vals derived from	
XX	S. pneumoniae.	
XX		
SQ	Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;	
	Alignment Scores:	
	Pred. No.: 0.0223 Length: 2652	
	Score: 130.00 Matches: 106	
	Percent Similarity: 32.81% Conservative: 62	
	Best Local Similarity: 20.70% Mismatches: 166	
	Query Match: 3.50% Indels: 178	
	DB: 24 Gaps: 25	
US	10-008-355-2 (1-712) x ABK52414 (1-2652)	
QY	192 AspValPhe-----LysAspValArgMetValPheAlaPro 203	
Db	T	T
	76 GATGTTTTCAAGCCTTCAGCGGATCAAAAGCGTAGCGCTATTCAATCGTTATTCACCA 135	
QY	204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223	
Db	T	T
	136 CCAACGTTACAGGTAACTT-----CACCTT 162	
QY	224 GlyAspPheSerValPheArgValTyrLagGlyAlaAspAsnArgProAlaGluTyrSer 243	
Db	T	T
	163 GGT-----CACGCT 171	
QY	244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263	
Db	T	T
	172 TGGGATACA---ACTTTGCCAGATATTATCATCCGTCAAAACGATGCACGAGTTTGTAT 228	
QY	264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282	
Db	T	T
	229 -----ACCTTTGGCTTCTCGGATGGACCACGACGCGGATTCGCACT 270	
QY	283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302	
	T	T





```

PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
DR DR P-PSDB; AAU38090.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 9586; 51pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2652 BP; 736 A; 602 C; 621 G; 693 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.0433 Length: 2652
Score: 127.00 Matches: 105
Percent Similarity: 33.01% Conservative: 64
Best Local Similarity: 20.51% Indels: 165
Query Match: 3.41% Gaps: 25
DB: 23
US-10-008-355-2 (1-712) x AAS55949 (1-2652)
Qy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
Db 76 GATGTTTCAAGCGCTTCAGGGCATCAAAGGCTAAGCGTTATTCAATCGTGATTCACCA 135
Qy 204 ProSerValGlyLySPheGlyLySPThrAspAsnTrpMetTrpProArgHisThr 223
Db 136 CCAACGTAAGTGGGAACGT-----CACGCTT 162
Qy 224 GlyAspPheSerValPheArgValTyraGlyAlaAspAsnArgProAlaGluTySer 243
Db 163 GGT-----CAGGCT 171
Qy 244 LysAspAsnLysProTyrLysProValTyraPheAlaValSerMetGlnGlyTyrrLys 263
Db 172 TGGGATACG---ACTTTCAGGATATCATCGTCAAAAACGCATGCAGGGCTTTGAT 228
Qy 264 AlaaspaspyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrrLeu---Thr 282
Db 229 -----ACCCTTTGCCGTTCCAGGTATGGACCACGCTGGTGTATTCGCACT 270
Qy 283 SerTrpGlyValGluAspArqIleGluAsnGluAspAsnProArqIleGluValArgGly 302

```



```
Db 271 CAGGCTAAGGTTGAGGACGCTTACGTTGGTGGGTATATACGCGTTATGACCTAGGTCGT 330
Qy 303 -----IleLysGlnGlyIleThrLysGluAlaMetSerAlaAspGlnAlaThr 318
Db 331 GAGTCTTCTTGGACGAAGCTCGGAATGGAAGAC-----GATATGCCACT 378
Qy 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
Db 379 ACCATCAAGGAACAAATGGGCAAGATGGGCTTCTGTAGACTATCTCGTGAGGTTTC 438
Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
Db 439 ACTCTGTGACGAAGGTTGTCA-----AAAGCTGTCGTAAG 474
Qy 358 AlaPheAlaAspTyrIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
Db 475 GTCCTTTGGACCTTACAAAGAGGCTGATCTATCTGCTGAGTTTATCATCAACTGG 534
Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
Db 535 GACCCAGCAGCTCGCACAGCCCTTCTGATATTGAGGTGATTCACAAGGATGAGAAGGT 594
Qy 395 -----GluGlyAlaLysAlaAsnArgGluMetThr 394
Db 595 GCCTTCTACCATGAATACATGCTGGAAGATGCTCACGCGCCCTTGAAGTTGCTACA 654
Qy 395 TyrLeuSerGluThrLeuPheGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
Db 655 ACTCTCTGAGACTATGTTGGGACGTT----- 684
Qy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434
Db 685 GCGGTGCGGTGTTAATCCAGAA-----GACCCGCGCTAC 717
Qy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 718 AAGGACTTGATTGTTAAACGCTCATCTTCCA---ATCGCTAATAAACTCATCCCAATC 774
Qy 449 MetLeuAsp-----IleValArgArg 455
Db 775 GTTGAGATGACGACGAGATCCTGAGTTGGTACTGGTGTGTAATAATCACACTGCC 834
Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
Db 835 CACGATCAAAATGACTTCTTGTTGGCCCAAGCTCATAACTTCCCAAGTCAACGTCATG 894
Qy 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db 895 AACGACGACGGAACATGATGAATGAGCTTGCCTTGAATTTTCAGGCATG-----GAT 945
Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
Db 946 CGTTTGTGAGCTCGT-----AGGCAGCTGCTGTCGTGAGTTGGAAGAATATC 990
Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaArgAla----- 531
Db 991 GGTGCCCTTGTCAAAATCGAAAACGTTGCCATTTCAGTCGGTCACTCAGACGGACAGGT 1050
Qy 532 -----IleGlnAla 534
Db 1051 GTCGTAGTTGAGCCACGCTTGTCTACTCAATGGTTGCTCAAGATGGACCAATGGCTAAG 1110
Qy 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
Db 1111 AACGCCATTGCCAACACACAGACAGAGACAAG----- 1143
Qy 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565
Db 1144 GTCGAATCTACCCACCTGTTTCAACGATACCTTCTTCAATGATGGAATATGCCAC 1203
Qy 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
```

```
Db 1204 GACTGGGTATCTCTCGTCAGCTTTGGTGGGTCCAC-----CAAAATC 1245
Qy 586 GlyAlaTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
Db 1246 CTGCTGGTGCACAT-----GCTGATGCTGAATGTATGTCGGCGAAGAAGCTCCAGAA 1299
Qy 606 SerAspGluPheAlaValGlnGlnGluAsnIleLeuAsp 617
Db 1300 GGTGACGATGACTCTCAGGACGACGACGCTCTGGAT 1335

RESULT 12
ABN68148
ID ABN68148 standard; DNA; 4110 BP.
XX AC ABN68148;
XX DT 01-JUL-2002 (first entry)
XX DE Streptococcus polynucleotide SEQ ID NO 4209.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J. Masignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR P-PSDB; ABP27517.
XX PT New Streptococcus protein for the treatment or prevention of infection
XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX PT for detecting a compound that binds to the protein -
XX PS Claim 7; Page 3577; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins.
SQ Sequence 4110 BP; 1462 A; 620 C; 791 G; 1237 T; 0 other;

Alignment Scores:
Pred. No.: 0.087 Length: 4110
```

Score:	126.50	Matches:	163
Percent Similarity:	33.29%	Conservative:	111
Best Local Similarity:	19.81%	Mismatches:	260
Query Match:	3.40%	Indels:	289
DB:	24	Gaps:	43
US-10-008-355-2 (1-712) x ABN68148 (1-4110)			
Qy	35	AsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeu	54
		:	
Db	1831	AATCCAGATAATCAGTCTATTCTGGAGAGATATCGCCAAACTCTAACATATTTTCAAGAC	1890
Qy	55	TyrSerPheAspLysProSerIleLaAsnAlaValIlePheGlyGlyCysThr	74
		:	
Db	1891	AGAGAATGATTAAAGAGCGCTCTGAAACTATAAGATCTTTT	1935
Qy	75	GlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAla	94
		:	
Db	1936	-----ACAGAGTCACAACTAAAAAACTCTATCGCTGCATCTATCGCTGGGGACGA	1989
Qy	95	Ile-----GlnSerGlnSerThrValAspHis	103
		::: ::	
Db	1990	TTGTCGCTGAAGTTAATCAATGGTATTCGAGATAAAGAGAGTCAAAAAACAATCTTCGAC	2049
Qy	104	AspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIleProGly	123
		:::       :	
Db	2050	TATCTTTATGATCGTAGATTAATCGCAACTTTATGCGAGTTGATTAATGATGATGGT	2109
Qy	124	LeuSerValLysTyrLeuArgLysIleValLysVal-----ThrAsp	137
		::: ::	
Db	2110	CTATCTTTTCAATCAATATATCAGTAAGCGACAGCGTGGTAGCTATTTCAGATAATCTAAAA	2169
Qy	138	LysValGluGlyGlnLeuLysGly-----	145
		:::           ::	
Db	2170	GAAGTTCTAGTGAGCTTCGAGTGACCCCTGCTATTAAAAAGGAAATTCACAAAGTTTG	2229
Qy	146	---IleThrAspGluMetGluArgLeuArg-----LysAlaGlnGluValCysGlnGlu	162
		:::       :	
Db	2230	AAAAATGTTGATGAGCTTCTTAAAGTCATGGGATACGAACCTGAACAAATTTGTGGTTGAG	2289
Qy	163	LeuAlaLysGluAsnAlaAspGluAsnGln-----	173
		:::       :	
Db	2290	ATGGCG---CGTGAGAATCAAAACAATCAAGGTCGCTGTAATCTCGACAAACGCTAT	2346
Qy	174	-----LeuCys-----IleValGlu	178
		::: ::    :	
Db	2347	AACTCTTCGATGATGGCGTTTAAAGATCTAGCTAGTACTGATGCAATATTTTGGAA	2406
Qy	179	ProPheTyrSerAsnAsnGlu-----TyrPhe-----	187
		GAATCTCTACGGATAATCAAGCGTTGCAAAATGAAGACTTTTCTCTTACTACTTACAA	2466
Db	2407	GAATATCTTACGGATAATCAAGCGTTGCAAAATGAAGACTTTTCTCTTACTACTTACAA	2466
Qy	188	-----LeuIleValTyrAsp	192
		:::     :	
Db	2467	AACGGAAGAGATATGATACAGGGGAAGCTCTAGATATTGACAATTTTAAAGTCAATATGAT	2526
Qy	193	ValPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPheGlyGly	212
		::: ::    :	
Db	2527	ATTGACCACATATCTCCTCAAGCTTTTCAAAAGATGATTTCTATT-----	2571
Qy	213	AspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyr	232
		::: ::	
Db	2572	-----GATAATCGTGT-----TTTGGTA	2589
Qy	233	AlaGlyAlaAspAsnArg-----ProAlaGluTyrSerLysAsp	245
		:::	
Db	2590	TCATCTGCTAAAAAATCGTGGAAAGTCAGATGATGTTCTCAGCTTGAATTTGAAAAAGAT	2649
Qy	246	AsnLysProTyr---LysProValTyrPheAlaAlaValSerMetGln-----	260
		::: ::	
Db	2650	TGTAAAGTTTCTGGAAAAAATTAATCTGATGCTAGTTAATGAGTCAAGCTAGTATGAT	2709
Qy	261	-----GlyTyrLysAlaAspAspTyrAlaMetThrIleGly	272

Db	2710	AATTGTGACTAAGCGAGCGCGGAGCCTAACTCCGATGATGAAGCAAGATTTATCCAA	2769				
Qy	273	PheProGlySerThrAspArgTyrLeuThrSerTrpGly	289			-----ValGluAspArg	289
Db	2770	CGTCAGTGGTGGAGACACACAATTTACCAAGCATGTTGCCGTATCTTGGATGAACGC	2829				
Qy	290	IleGluAsnGluAsnPro	303			-----ArgIleGluValArgGlyLeu	303
Db	2830	TTTAATATGAGCTTGATAGTAAGGTAGAGGATCCGCAAAAGTTAAATTTGTAACCTTG	2889				
Qy	304	LysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAla	323			-----ValGluAspArg	323
Db	2890	AAAGTCAAAATTTGGTTTCCAAATTTCCGAAAGAATTTGGATCTATATAAATTCGTGAAGTT	2949				
Qy	324	SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeu	343			-----ValGluAspArg	343
Db	2950	AAACAATATCACCATCGACATGCTATCTTAATGACAGTAGTTGCTFAAAGCTATTCTCA	3009			-----ValGluAspArg	3009
Qy	344	AlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIle	363			-----ValGluAspArg	363
Db	3010	ACCAAAATCCTCAGTTA	3039			-----ValGluAspArg	3039
Qy	364	ArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyr	383			-----ValGluAspArg	383
Db	3040	-----GTCTACGCGCATAT	3072			-----ValGluAspArg	3072
Qy	384	LysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGlu	398			-----ValGluAspArg	398
Db	3073	AAAAACGGTAAATCCCGCTACAGAAAAGCTATTTTCTATTCAAAATATTATGAACCTCTTT	3132			-----ValGluAspArg	3132
Qy	399	-----ThrLeuPheGlyThrGluValValArgPheAlaGlnPheAlaAsn	414			-----ValGluAspArg	414
Db	3133	AAACTAAGGTAACTTTAGCGGATGGAACCGTTGTTGTAATA	3177			-----ValGluAspArg	3177
Qy	415	AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr	434			-----ValGluAspArg	434
Db	3178	GATATTCAAGTTAATAATGATACGGGT	3225			-----ValGluAspArg	3225
Qy	435	LysAspTyrLeuProSerLeuAspArgLysValLeu	452			-----ValGluAspArg	452
Db	3226	AAACACTTT	3279			-----ValGluAspArg	3279
Qy	453	ValArgArgArgIlePro	458			-----ValGluAspArg	458
Db	3280	AAGAAGACAGATTACAGACAGTGGTTCTCTAGGAATCAATCTGGCCCATGTTAAC	3339			-----ValGluAspArg	3339
Qy	459	AlaAspLysLeuProAspIlePheTyrAsnValIleAspLysLysPheGly	476			-----ValGluAspArg	476
Db	3340	TCAGATAAGTTG	3378			-----ValGluAspArg	3378
Qy	477	---AspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp	495			-----ValGluAspArg	495
Db	3379	TTAGATCCTAAGAAATATCGAGGT	3432			-----ValGluAspArg	3432
Qy	496	LysPheHisAlaMetLeuLysSer	503			-----ValGluAspArg	503
Db	3433	TTAGTTCTAGCTGATATCAAAAAGGTTAAGCACAAAAACTAAACACAGTTACCGAACATT	3492			-----ValGluAspArg	3492
Qy	504	-----MetAspLysGluLysPheAlaLysAlaIleGluLysAspProAla	518			-----ValGluAspArg	518
Db	3493	TTAGGAATTAACCATCATCGAGAGGTCACAGATT	3540			-----ValGluAspArg	3540
Qy	519	ValGluLeu---SerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMet	537			-----ValGluAspArg	537
Db	3541	GCTTTCCCTTGAATCAAAAGGCTATTTA	3588			-----ValGluAspArg	3588
Qy	538	Ala-----AsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAla	552			-----ValGluAspArg	552
Db	3589	ATTATTTTGGCCCAAGTATAGTCTGTTCGAATTTAGAAAAATGGCGCTGCTGCTACTTGTCT	3648			-----ValGluAspArg	3648
Qy	553	GlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThr	570			-----ValGluAspArg	570

Db 3649 AGTCTGGTGAATTCACAAAAGGTAATCAGCTAGCTTACCAACACAAATTTATGAAGTTC 3708  
 Qy 571 -----MetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp----- 585  
 Db 3709 TTATACCTTGCAAGTCGTTATTAAGTCAAAAGGT---AAACAGAGGAGATTGAGAG 3765  
 Qy 586 ---GlyAlaTrpTyrAsnTyrHisThr----- 594  
 Db 3766 AACCAAGAATTTGTAATCAACATGCTCTTATTTTGTGATGACATCCTTCAATTAATTAAT 3825  
 Qy 595 -----GlyLysGlyVal-----LeuGluLys----- 601  
 Db 3826 GATTTTCAAAACAGGTATTCTAGCAGATGCTAATTTAGAGAAAATCAATPAAGCTTTAC 3885  
 Qy 602 GlnAspProLysSer-----AspGluPheAlaValGlnGluAsnIleLeuAsp 617  
 Db 3886 CAAGATAATAGGAATATATATCATCATGATCACTGCT-----AATAATATTATCAAT 3939  
 Qy 618 LeuPheArgThrLysAsnTyrGly----- 625  
 Db 3940 CTATTTTACTTTTACAGCTAGGAGCTCCAGCAGCTTTTAAATTTTGTATAAAATAGTT 3999  
 Qy 626 -----ArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsp 642  
 Db 4000 GATAGAAACCGCTATACATCAACTAAAGAAAGTACTTAATCTCACCCTAATTCATCAATCT 4059  
 Qy 643 IleThrGly 645  
 Db 4060 ATTACTGGA 4068  
 RESULT 13  
 AAZ31948  
 ID AAZ31948 standard; DNA; 2142 BP.  
 AC AAZ31948;  
 XX  
 XX 26-JAN-2000 (first entry)  
 XX  
 XX M. catarrhalis strain LES1 tbpB gene.  
 XX  
 KW TbpB gene; Tbp2; transferrin binding protein; diagnosis; otitis media;  
 KW genetic immunisation; Moraxella infection; antigen; vaccine; detection;  
 KW antitumour antibody production; therapy; ss.  
 XX  
 XX Moraxella catarrhalis.  
 XX  
 XX WO9552947-A2.  
 XX  
 XX 21-OCT-1999.  
 XX  
 XX 12-APR-1999; 99WO-CA00307.  
 XX  
 XX 14-APR-1998; 98US-0059584.  
 XX  
 XX (CONN-) CONNAUGHT LAB LTD.  
 XX  
 XX Myers LE, Schryvers AB, Harkness RE, Loosmore SM, Du R, Yang Y;  
 PI Klein MH;  
 XX  
 XX WPI; 1999-620376/53.  
 DR P-PSDB; AAY43380.  
 XX  
 XX Nucleic acid encoding transferrin binding protein 2 of Moraxella  
 PT catarrhalis, useful for diagnostics, immunization and recombinant  
 PT protein production  
 XX  
 XX Claim 2; Fig 6; 114pp; English.  
 XX  
 XX This sequence encodes the Moraxella catarrhalis strain LES1 transferrin  
 CC binding protein (Tbp2) of the invention. This sequence is also referred  
 CC to as the TbpB gene. The TbpB gene is used to produce recombinant Tbp2;  
 CC for identification or diagnosis of Moraxella, or for cloning related  
 CC species, using hybridisation assays; and for genetic immunisation against

CC Moraxella infections, e.g. otitis media. The Tbp2 proteins are useful as  
 CC antigens, either in vaccines (including components of conjugate vaccines  
 CC that contain antigens from other bacteria or from tumours, in which case  
 CC they elicit production of antitumour antibodies that may be coupled to  
 CC chemotherapeutic agents or biologically active agents) or to raise  
 CC antibodies (for use as diagnostic reagents and for treating Moraxella  
 CC infections), also for detecting Moraxella antibodies.  
 XX  
 SQ Sequence 2142 BP; 753 A; 458 C; 457 G; 474 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.0705 Length: 2142  
 Score: 123.50 Matches: 125  
 Percent Similarity: 32.17% Conservative: 77  
 Best Local Similarity: 19.90% Mismatches: 273  
 Query Match: 3.32% Indels: 153  
 DB: 20 Gaps: 25

US-10-008-355-2 (1-712) x AAZ31948 (1-2142)

Qy 119 LeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLys 138  
 Db 466 ATTCTTTTGTATAAATCTTATTGAATACCTTAAAAAATCATCCGAGTTGTAGTAAA 525  
 Qy 139 ValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGlu 158  
 Db 526 TTTGAAGCACAAAAAGCGGTATTGAAAATAACACAGACTGACACACAAAGATTATCA 585  
 Qy 159 ValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGlu 178  
 Db 586 TCAGAGCAAAAAGAGCAAAAGTCAAGAGCGTTGGACAATGCTTTA----- 633  
 Qy 179 ProPheTyrSerAsnAsnGluTyrPheLeuIleVal-----TyrAspValPheLys 195  
 Db 634 ACTCAATTTGCCAAGAAAATAACAGGAGCTAATTGAGACGCCCAATGATAAAATCT 693  
 Qy 196 AspValArgMetValPheAlaProProSerSerValGlyLysPheGlyAspThrAsp 215  
 Db 694 GACGCACGC----- 702  
 Qy 216 AsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAla 235  
 Db 702 ----- 702  
 Qy 236 AspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAla 255  
 Db 703 ---AACCGTGATCTAGAATATGTCAAAGTCTGGTTTAACTATCTTTCTGGATATACCGCC 759  
 Qy 256 Ala-----ValSerMetGlnGly-----TyrLysAla 264  
 Db 760 ACCGACACGACAAAAACCAATATCGTGCCTATTATGTCGCTTCTATATAAAGGC 819  
 Qy 265 AspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrp 284  
 Db 820 AGCGAAACCGCCAAAGAGCTA-----CCACAACACAGTCGCAAAATATAAAGGTATTGG 873  
 Qy 285 Gly-----ValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302  
 Db 874 GACTTTATGACAGATGCGCACCATTTGATAACAAATACACG-----GATTGCCAGGT 924  
 Qy 303 Ile---LysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln---AlaThrArgIle 320  
 Db 925 ATGCCACACAAACCCAGTGGCTAGTCTTGTCTACTGATGATGATGATGATGATGATG 984  
 Qy 321 -----LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332  
 Db 985 ACAGACAAAAAATAACAGCCCGCAGTGATTACAATGTCATATGTCATAGCAGTGAATTT 1044  
 Qy 333 TrpLysAsnSerIleGly-----MetAsnArgGlyLeuAlaArg 345  
 Db 1045 GATGTTAAATTTGCTGATAAAAAAATTAAAGCAACCTTATCAGTAATCAGTATTCAGGC 1104  
 Qy 346 LeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLys 365

```
Db 1105 ACAGCTGTAACGCGCAAGAGCGTTATATAAATAGAACTGATATCCACGCGCAACCGCTTC 1164
Qy 366 AsnGlyLysSerAlaValTyrGlyAspValLeuSerLeuGluLysAlaTyrLysGlu 385
Db 1165 CGTGCACTGCC-----ACCGCAAGCGATAAAGCA-----GAA 1197
Qy 386 GlyAlaLysAlaAsnArgGluMetThrTyr---LeuSerGluThrLeuPheGlyGlyThr 404
Db 1198 GACAGCAAAACCCCAACACCCCTTTACCAGCGATGCTACAACAGCTAGAGGTGTTT 1257
Qy 405 GluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla 424
Db 1258 TATGCAACCAAGGCGAGGAGCTGCGAGTAAATTTCTTAACCGATGACAACAACACTCTT 1317
Qy 425 GlyLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAspArgLys 444
Db 1318 GGGGTCTTGGTGTAAACGAGATAGATAAAGCAACCGAGCGCATCTTAGATGCCTAT 1377
Qy 445 ValLeuProAlaMetLeuAspIleValArgArgIleProAlaAspLysLeuProAsp 464
Db 1378 GCACTT-----GGG 1386
Qy 465 IlePheLysAsnValIleAsp---LysLysPheLysGlyAspThrLysLysTyrAlaAsp 483
Db 1387 ACATTTAATAATACAAATAAAGCAACCACTTACCCCATTTACCAAAAAACAACCTGGAT 1446
Qy 484 ---PheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
Db 1447 AACTTTGGCAATGCCAAAGTGTGTGGTCTACCGTCATTAATTTGGTGTCTAC 1506
Qy 503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer 522
Db 1507 GATGCCACCAAAATGAATTCACCAAAAAATTCACCAAGAC----- 1548
Qy 523 LysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 542
Db 1549 AAGCAACTTCTGCCACCAAAAGCGGCGAGACT---TTGATGGTGAATGATCAAGTT 1605
Qy 543 IleGluLys-----GlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
Db 1606 ATCGTCAAAACCTATGGCAAAACCTTGAATACCTAAATTTGGTGACCTAGTGTCCGT 1665
Qy 561 ArgAla-----LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGly 576
Db 1666 GATAGCCATAGCGTCTTTTACAGGCGAAGCGACCGCTACCAAGCGGAGAAAGCGGTA 1725
Qy 577 SerIleLysGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHisThr----- 593
Db 1726 CCAACCAAGCAAGCAAGCAAAATATCTGGGGAACCTGGGTAGGATACATCACAGGAGCGGC 1785
Qy 594 ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGlu 613
Db 1786 ACAGCAAAACCTTAATAGGCCCAAGAT-----ATTGCTGTAT 1824
Qy 614 AsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeu 633
Db 1825 TTTGACATTGACTTTGAGAGAAATACAGTTAAAGCAAACTGACCAACCAAGCGCGACA 1884
Qy 634 HisIleAlaPheLeuSerAsnAspIleThrGly----- 645
Db 1885 GATCCTGTCTTTAATCAATCAAAAGGTGAATTCAGGCAATGGCTGGACAGCGCAAGCCAGC 1944
Qy 646 -----GlyAsnSerGlySerProVal 652
Db 1945 ACCACCAAGCGGAGGAGGCTACAAGATAGATTCTACCATGATCAGGCAAAATCCATC 2004
Qy 653 PheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSer 672
Db 2005 GTCATCGAAATGCGGAAGTTACTTGGGGGCTTTTATGTGTCCAAATGCAACAGCAGATGGC 2064
Qy 673 GlyAspIleGluPheProAsp 680
Db 1105 ACAGCTGTAACGCGCAAGAGCGTTATATAAATAGAACTGATATCCACGCGCAACCGCTTC 1164
Qy 366 AsnGlyLysSerAlaValTyrGlyAspValLeuSerLeuGluLysAlaTyrLysGlu 385
Db 1165 CGTGCACTGCC-----ACCGCAAGCGATAAAGCA-----GAA 1197
Qy 386 GlyAlaLysAlaAsnArgGluMetThrTyr---LeuSerGluThrLeuPheGlyGlyThr 404
Db 1198 GACAGCAAAACCCCAACACCCCTTTACCAGCGATGCTACAACAGCTAGAGGTGTTT 1257
Qy 405 GluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla 424
Db 1258 TATGCAACCAAGGCGAGGAGCTGCGAGTAAATTTCTTAACCGATGACAACAACACTCTT 1317
Qy 425 GlyLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAspArgLys 444
Db 1318 GGGGTCTTGGTGTAAACGAGATAGATAAAGCAACCGAGCGCATCTTAGATGCCTAT 1377
Qy 445 ValLeuProAlaMetLeuAspIleValArgArgIleProAlaAspLysLeuProAsp 464
Db 1378 GCACTT-----GGG 1386
Qy 465 IlePheLysAsnValIleAsp---LysLysPheLysGlyAspThrLysLysTyrAlaAsp 483
Db 1387 ACATTTAATAATACAAATAAAGCAACCACTTACCCCATTTACCAAAAAACAACCTGGAT 1446
Qy 484 ---PheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
Db 1447 AACTTTGGCAATGCCAAAGTGTGTGGTCTACCGTCATTAATTTGGTGTCTAC 1506
Qy 503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer 522
Db 1507 GATGCCACCAAAATGAATTCACCAAAAAATTCACCAAGAC----- 1548
Qy 523 LysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 542
Db 1549 AAGCAACTTCTGCCACCAAAAGCGGCGAGACT---TTGATGGTGAATGATCAAGTT 1605
Qy 543 IleGluLys-----GlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
Db 1606 ATCGTCAAAACCTATGGCAAAACCTTGAATACCTAAATTTGGTGACCTAGTGTCCGT 1665
Qy 561 ArgAla-----LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGly 576
Db 1666 GATAGCCATAGCGTCTTTTACAGGCGAAGCGACCGCTACCAAGCGGAGAAAGCGGTA 1725
Qy 577 SerIleLysGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHisThr----- 593
Db 1726 CCAACCAAGCAAGCAAGCAAAATATCTGGGGAACCTGGGTAGGATACATCACAGGAGCGGC 1785
Qy 594 ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGlu 613
Db 1786 ACAGCAAAACCTTAATAGGCCCAAGAT-----ATTGCTGTAT 1824
Qy 614 AsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeu 633
Db 1825 TTTGACATTGACTTTGAGAGAAATACAGTTAAAGCAAACTGACCAACCAAGCGCGACA 1884
Qy 634 HisIleAlaPheLeuSerAsnAspIleThrGly----- 645
Db 1885 GATCCTGTCTTTAATCAATCAAAAGGTGAATTCAGGCAATGGCTGGACAGCGCAAGCCAGC 1944
Qy 646 -----GlyAsnSerGlySerProVal 652
Db 1945 ACCACCAAGCGGAGGAGGCTACAAGATAGATTCTACCATGATCAGGCAAAATCCATC 2004
Qy 653 PheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSer 672
Db 2005 GTCATCGAAATGCGGAAGTTACTTGGGGGCTTTTATGTGTCCAAATGCAACAGCAGATGGC 2064
Qy 673 GlyAspIleGluPheProAsp 680
```

```
Db 2065 GGGTCATTACACACGATACCCAT 2088
RESULT 14
ABQ69595
ID ABQ69595 standard; DNA; 4716 BP.
XX AC
XX ABQ69595;
DT 29-AUG-2002 (first entry)
XX XX
DE Listeria innocua DNA sequence #1034.
XX XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX XX
OS Listeria innocua.
XX XX
PN WO200228891-A2.
XX XX
PD 11-APR-2002.
XX XX
PF 04-OCT-2001; 2001WO-FR03061.
XX XX
PR 04-OCT-2000; 2000FR-0012697.
XX XX
PA (INSP ) INST PASTEUR.
XX XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX XX
PI Kunst F, Glaser P;
XX XX
DR WPI; 2002-332479/37.
XX XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators
XX XX
PS Claim 7; SEQ ID 2408; 180pp; French.
XX XX
CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 4716 BP; 1750 A; 734 C; 1110 G; 1122 T; 0 other;

Alignment Scores:
Pred. No.: 0.253 Length: 4716
Score: 122.50 Matches: 138
Percent Similarity: 32.84% Conservative: 104
Best Local Similarity: 18.72% Mismatches: 308
Query Match: 3.29% Indels: 187
DB: 24 Gaps: 31

US-10-008-355-2 (1-712) x ABQ69595 (1-4716)
Qy 11 GlyAlaAlaLeuLeuGlyAlaSerGlyValAlaLysAlaAsp----- 25
Db 1675 GCGCGAGTTGTTATTGTCCTGGGGGAAAAGGAAAGGAAAGTTAACTTTTCAGCAAGCT 1734
Qy 26 -----LysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArg 41
Db 1735 ACTCCAGGACTCAAGGATGGTAAATTCGGTAAGTTG-----GCTAGCGGT 1782
Qy 42 MetArgGluLeuGlyPheThrLeuProLeuAspSerPheAspLysProSer 61
Db 42 MetArgGluLeuGlyPheThrLeuProLeuAspSerPheAspLysProSer 61
```

Qy	360	AlaaspTpIleAArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeu	379
Db	2734	CAAAAAGAGTTAGAAAG-----	2757
Qy	380	GluLysAlaTyrLysGluGluAlaLysAlaAsnAArgGluMetThrTyrLeuSerGluThr	399
Db	2758	CAAAAGCAAAAGGAAGAGCTGCTATTGCACCTCCTCAGCTTAAGAACACAAAAATA	2817
Qy	400	LeupheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr---	418
Db	2818	ATACAGCAAACTTAGAAAACTCGAAAGAAAAAATGTCAGCAAAAGCAGCAGCAAGTGT	2877
Qy	419	-----AsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLys	435
Db	2878	GTTAAAAATTCCGAAAAACCTAGACATGCATCCGTAAGAAGAGCTAATAAAGACATATAA	2937
Qy	436	AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArg	455
Db	2938	CAGACTAAAAAATACTTGATGAAAA-----CGATTTGTCACAGT	2979
Qy	456	ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys	475
Db	2980	GAATTTCCAGAGAAGAATATCAAGATCGCGTGAAAAATGCG-----AAGAAAAAGAAA	3033
Qy	476	GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp	495
Db	3034	AACGCGCTAGTTAAAGAGCGAGAGATGCACGATAATGTATGTAGTCGAGAAGCTCAGCAA	3093
Qy	496	LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys	515
Db	3094	CAAGCAAAAGGACATTTAAGAGAAGTAGATCGGGAACACT--GGTGAACACATTAAATAAG	3150
Qy	516	AspProAlaValIleLeuSerLysSerValIleAlaAlaAArgAlaIleGlnAlaAsp	535
Db	3151	-----TGGAGCAATTTAAAAAAGGACACAGTTGAACATTTAAACAGTATAAAGATGCA	3204
Qy	536	AlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArg	555
Db	3205	CGCCTCGGAAAATGGAACGAGTTGTCCTCGCGACAGTTAATATTTCAGTGGAAATGAAA	3264
Qy	556	GluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyr	575
Db	3265	GATACA---CGCTTCAGAGTATTTCAATAGTTTAAAGACAACTTGTATAAAGTCATTAT	3321
Qy	576	GlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThr-----	593
Db	3322	AATGTAATACCGGAGTCAAT--AAAGTACTATCGTCTCTTAAACATTAAAGACAATACCT	3378
Qy	594	---ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln	612
Db	3379	GTGCTTGGGAACGGTAGCATGGTTCAAGT-----CAA	3411
Qy	613	GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGln	632
Db	3412	GAACAACAGTTTATCTGCAAGAGATGAAGAAACGTTATCATCTCACATCCCAATCAGGAAT	3471
Qy	633	LeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly-----GlyAsnSer	648
Db	3472	CTAGCTATGAACTATACTGGTCTTAATATGCAAGTGGCCAAATATTGCTGTGGAGAA	3531
Qy	649	GlySerProVal---PheAspLysAsn-----GlyArgLeuIleGly	661
Db	3532	GGTTTCGAGATTGCATATAACAAAGAACTCAGCAACAGCGAGAAATCTCTAGG	3582

DT 29-AUG-2002 (first entry)

DE Listeria innocua DNA sequence #675.

XX Antibacterial; Listeria; food contamination; mutational analysis;  
KW infection; ds.

XX Listeria innocua.

XX W0200228891-A2.

XX 11-APR-2002.

XX 04-OCT-2001; 2001WO-FR03061.

XX 04-OCT-2000; 2000FR-0012697.

XX (INSP ) INST PASTEUR.

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX Kunst F, Glaser P;

XX WPI; 2002-332479/37.

XX New genomic sequences from Listeria species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators -

XX Claim 7; SEQ ID 686; 180bp; French.

XX The present invention relates to nucleic acid sequences  
CC (AB067188-AB067192) from Listeria sp. The sequences are useful as probes  
CC and primers for identification and/or detection of Listeria (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of Listeria (potential therapeutic agents), also for  
CC treating infections by Listeria, and are useful as immunogens in  
CC anti-Listeria vaccines.

CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 4731 BP; 1752 A; 735 C; 1117 G; 1127 T; 0 other;

Alignment Scores:

Pred. No.:	0.254	Length:	4731
Score:	122.50	Matches:	138
Percent Similarity:	32.84%	Conservative:	104
Best Local Similarity:	18.72%	Mismatches:	308
Query Match:	3.29%	Indels:	187
DB:	24	Gaps:	31

US-10-008-355-2 (1-712) x AB067873 (1-4731)

Qy 11 GlyAlaAlaLeuLeuLeuGlyAlaSerGlyValAlaLysAlaAsp----- 25  
Db 1690 GCGGAGTGTGATTATTCCTGGGGGAAAAAGGAAAGGAAAGTTAAACTTTTCAGCAAAAGCT 1749  
Qy 26 -----LysGlyMetTrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArg 41  
Db 1750 ACTCAGCACTCAAGCATGGGTAAATTCGGTAAGTTG-----GCTAGGCGGT 1797  
Qy 42 MetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysProSer 61  
Db 1798 GTAGTGTTTTAGGTGTAGGATTAAGCGCTACTCAGTTGATTGATGATGACAAAAATAAT 1857  
Qy 62 IleAlaAsnAlaValAlaIlePheGlyGlyCysThrGlyIleThrValSerAspGln 81  
Db 1858 GCAGGTGAAAAAACTGGTGGCTTTGGTGGTCTTTAGTGCGT-----ATGGCTGGTGGT 1911  
Qy 82 GlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAla----- 94  
Db 1912 GCTCATAGGAACATGATTCCTGGAGTAGCTACAGCTATTGGTGGGCAATTGGC 1971

Qy 95 -----IleGlnSerGlnSerThr 100  
Db 1972 GCTTTTGTGCACTGCATTAGGTAAAGAAATAGGAAATATGTTCAAAAAGAGGACCA 2031  
Qy 101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThr---MetGlyGluGluLeu 119  
Db 2032 AAAATATTAGATAAATTCAGACAGAGTTGGAAAGGCTTAAGCAAAATTTGCTGAAGAGCAT 2091  
Qy 120 ProIleProGlyLeuSerValTyrLeuArgLysIleValLys----- 134  
Db 2092 CCGATTTTAGGAGCAAAATATTAATGTAATAAATAAACTATCGATCGGCGCAAAAGAGGC 2151  
Qy 135 -----ValThr 136  
Db 2152 ATTAAGCAGTTGGAGACACGCATAAACTGTTTGAATGCTTCAAAATCAATAGTCGCC 2211  
Qy 137 Asp-----LysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArg 154  
Db 2212 GACCGCTCAAAATTTGATCGGAGCAGTAAGGTGTTTCTAAAGAC----- 2256  
Qy 155 LysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeu 174  
Db 2257 TCTGCTAAAGCAATGAACGAGTATTTAGGAATGAGCAGAAAAATGCAAGATAGTCGAGTC 2316  
Qy 175 -----CysIleValGluProPheTyrSerAsnGluTyrPhe 187  
Db 2317 GAANTCATGGTATCTGTCGTCGTATTACTCAAAAAGAACTCGAAAAACAAC----- 2367  
Qy 188 LeuIleValTyrAspValPheLysAspValArgMetValPheAlaProProSerSerVal 207  
Db 2368 ATAAAAACATATGATAAATATGCTCGAT---CAGTTAATAGTGTCAACTGAAAAAATCA 2424  
Qy 208 GlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSer 227  
Db 2425 GCAAAACCAATAAAGAT-----TGGGATAAGCTAGTAGCACTTTGGGGCT 2469  
Qy 228 ValPheArgValTyrAlaGlyAla-----AspAsnArgProAlaGluTyrSer 243  
Db 2470 GTTCTTAAAGATACAGCTAGCGCAAGAAACACACACTAATAATGAACCTGCATCAATAAT 2529  
Qy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263  
Db 2530 AAACGGGATATTAAAGCGCAATTAACGAAAGAAATTAAGCTTTAGAAAGAGAAATTCATATAA 2589  
Qy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSer 283  
Db 2589 ----- 2589  
Qy 284 TrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGlyIle 303  
Db 2590 ---GGCAAGAAGAGATTACACGAAATGCTGAAAT-----AGATTAAATGCTATT 2637  
Qy 304 LysGlnGlyIleTrpLysGlu-----AlaMetSerAlaAspGlnAlaThrArgIleLys 321  
Db 2638 AAACAGAAAAGCTACCAAGAAACGGAATTAACAAAGATGAAGAAAAAAGAAATCCAA 2697  
Qy 322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArg 341  
Db 2697 ----- 2697  
Qy 342 GlyLeuAlaArgLeuAsp-----ValIleGlyArgLysArgAlaGluArgAlaPhe 359  
Db 2698 -----AGAATGGAAACAAACACACTTGAGTTCCGCTGCTTCTGAACGACGTTCTTAT 2748  
Qy 360 AlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeu 379  
Db 2749 GAAAAAGCAAGCTAGAAAG----- 2772  
Qy 380 GluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThr 399  
Db 2773 GAAAGCAAGAAAGAAAGCTGCTATTGCACTCACTGCTTCAAGTAAAGCAAAATAA 2832  
Qy 400 LeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr--- 418

```
Db 2833 ATACTAGAAACTTAGAAACTCGAAAAGAAAAATGTCAGCAAAAGCAGCAGCAAGTGT 2892
Qy 419 -----AsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLys 435
Db 2893 GTTAAAAATCCGCAAAAGCTAGAGTGCATCCGTAAGAGCTAATAAGAGATATAA 2952
Qy 436 AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArg 455
Db 2953 CAGACTAAAAAATACTGTGATAAAA-----CGATTGTGCACAGGT 2994
Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
Db 2995 GAAATTTTCAGAAGAAGAAATATCAAGATGCGCTGAAAAATCGG-----AAGAAAAAGAAA 3048
Qy 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db 3049 AACGGCGTAGTTAAGAGCGGAGAGAGATGCACCATATGTAGTCGAGAGCTCAGCAA 3108
Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGlyLysPheAlaLysAlaIleGluLys 515
Db 3109 CAAGCAAAAGGACATTTTAAGAGAAGATAGATTGGAAACT---GGTGAACATTAATAAG 3165
Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaArgAlaIleGlnAlaAsp 535
Db 3166 -----TGGGAGCAATTTAAAAAGGACAGTTCGAACATTTAACAGTATATAAGATGCA 3219
Qy 536 AlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArg 555
Db 3220 CGCTCGGAANAATGGACGAGTTGTCTCTCGGACAGTATATATTTTCAGTGGAGTAAAA 3279
Qy 556 GluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyr 575
Db 3280 GATACA---GCGTTGAGAGTATTTGATAGTTTAAAGACAAACTTGTATAAAGTCATTAAT 3336
Qy 576 GlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThr----- 593
Db 3337 AATGTAATTACCGGAGTCAAT---AAAGTACTATCGTCTTTTAACATTAAGACAAATACCT 3393
Qy 594 ---ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 612
Db 3394 GTGCTTGGGACGCTAGCATAGTTCAGT-----CAAA-----CAA 3426
Qy 613 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGln 632
Db 3427 GAAAAACAGTTTATCTGCAAGAGGATAAGAAAACGTATCATCTACATCCCAATCAGGAAAT 3486
Qy 633 LeuHisIleAlaPheLeuSerAsnAsnAspIleThrGly-----GlyAsnSer 648
Db 3487 CTAGCTATGAACATATCTGGTCTTAATAAGTGCAGTGGCCCAAAATATGGCTGGTGAAGAA 3546
Qy 649 GlySerProVal---PheAspLysAsn-----GlyArgLeuIleGly 661
Db 3547 GGTTCGAGATTGCATATACAGAACTCAGCACACAGCGAGAAATTCAGGG 3597
```

Search completed: May 23, 2003, 10:05:47  
Job time : 379.182 secs





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 09:42:03 ; Search time 60.7132 Seconds  
(without alignments)  
3596.482 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MQMKLSILGALLLGASG.....LFMDKMGQCPRLIQELKLI 712

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPRO\_pool/US10008355/runat\_16052003\_110403\_2582/app\_query.fasta\_1.1230  
-DB=Issued\_Patents\_NA -QMT=fastcap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0.1  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10008355 -CGN1\_1\_47 -runat\_16052003\_110403\_2582 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_WMAP -LARGSEQURY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2995	80.5	1974	4	US-09-221-017B-726
2	743.5	20.0	1317	4	US-09-221-017B-382
3	441.5	11.9	2384	4	US-09-221-017B-1045
4	265.5	7.1	561	4	US-09-221-017B-12
5	130	3.5	2852	2	US-08-953-492-1
6	129	3.5	6004	4	US-08-961-527-27
c 7	123.5	3.3	2142	4	US-09-059-584-52
8	115	3.1	4853	2	US-08-793-824-1
c 9	111	3.0	4411529	4	US-09-103-840A-1
10	110.5	3.0	12665	4	US-08-961-527-134
11	110	3.0	2319	4	US-09-134-078-14
12	110	3.0	2667	4	US-09-134-001C-195

13	110	3.0	3279	4	US-08-446-137B-1	Sequence 1, Appli
c 14	109	2.9	15363	4	US-08-961-527-139	Sequence 139, App
c 15	109	2.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 16	107.5	2.9	4851	4	US-09-221-017B-999	Sequence 999, App
c 17	107	2.9	38155	4	US-09-453-702B-79	Sequence 79, Appl
c 18	106.5	2.9	15213	4	US-08-961-527-26	Sequence 26, Appl
19	106	2.9	2472	3	US-08-335-844A-7	Sequence 7, Appli
20	106	2.9	3358	3	US-08-335-844A-20	Sequence 20, Appl
21	106	2.9	3698	4	US-09-327-536-1	Sequence 1, Appli
22	105.5	2.8	1994	3	US-08-600-982-22	Sequence 22, Appl
23	105.5	2.8	1994	5	PCT-US94-10261A-22	Sequence 22, Appl
24	104	2.8	3731	4	US-08-811-583-1	Sequence 1, Appli
25	102.5	2.8	2571	4	US-09-134-001C-2601	Sequence 2601, Ap
26	102.5	2.8	6142	4	US-09-514-302-1	Sequence 1, Appli
27	102	2.7	1242	4	US-09-134-001C-2274	Sequence 2274, Ap
28	102	2.7	1415	4	US-09-345-214-11	Sequence 11, Appl
29	102	2.7	2008	4	US-09-345-214-12	Sequence 12, Appl
30	102	2.7	2491	2	US-09-345-214-5	Sequence 5, Appli
31	102	2.7	2582	2	US-08-816-105A-2	Sequence 2, Appli
32	102	2.7	4310	3	US-09-008-172-1	Sequence 1, Appli
33	102	2.7	4310	4	US-09-210-361-5	Sequence 5, Appli
34	102	2.7	4310	4	US-09-740-274-5	Sequence 5, Appli
35	101.5	2.7	2253	1	US-07-854-596B-39	Sequence 39, Appl
36	101.5	2.7	3727	1	US-08-249-380-1	Sequence 1, Appli
37	101	2.7	3668	4	US-09-302-620B-89	Sequence 89, Appl
38	100.5	2.7	2836	4	US-09-157-257-3	Sequence 3, Appli
39	100.5	2.7	3278	1	US-08-484-105-13	Sequence 13, Appl
40	100.5	2.7	3278	1	US-08-484-106-13	Sequence 13, Appl
41	100.5	2.7	5020	4	US-08-961-527-142	Sequence 142, App
42	100.5	2.7	8501	4	US-09-298-367B-6	Sequence 6, Appli
43	100	2.7	4330	3	US-09-310-293-1	Sequence 1, Appli
44	100	2.7	4330	4	US-09-579-376-1	Sequence 1, Appli
45	100	2.7	6744	1	US-08-119-125A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-221-017B-726

; Sequence 726, Application US/09221017B

; Patent No. 6444799

; GENERAL INFORMATION:

; APPLICANT: ROSS, Bruce C.

; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

; NUMBER OF SEQUENCES: 1120

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER

; STREET: 755 PAGE MILL ROAD

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1018

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSEQ for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/221-017B

; FILING DATE: 23-DEC-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1182

; FILING DATE: 31-DEC-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP1546

; FILING DATE: 30-JAN-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PP2911

; FILING DATE: 09-APR-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/AU98/01023

; FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:  
NAME: MONROY, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 726:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1974 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...1974  
US-09-221-017B-726

Alignment Scores:  
Pred. No.: 1.3e-317 Length: 1974  
Score: 2995.00 Matches: 576  
Percent Similarity: 99.65% Conservative: 0  
Best Local Similarity: 99.65% Mismatches: 0  
Query Match: 80.53% Indels: 2  
Dbs: 4 Gaps: 0

US-10-008-355-2 (1-712) x US-09-221-017B-726 (1-1974)

QY 137 AspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAla 156  
Db 1 GACAGGTGAGAGGACAGCTCAAGGGTATCTACAGAGATGGAGCGCTGCGCAAGCT 60  
QY 157 GlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIle 176  
Db 61 CAGGAGGTATGCCAAGACTGCCCAAAAGAAATGATGAGAGGAGCAACCAACTCTGCATC 120  
QY 177 ValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAsp 196  
Db 121 GTAGAGCCTTCTATTATCAACAACCAAGAACTCTCTCATCGTCTACGATGTATTCAAGAC 180  
QY 197 ValArgMetValPheAlaProProSerValGlyLysPheGlyGlyAspThrAspAsn 216  
Db 181 GTTCGTATGGTATTGCTCTCCAGCTCTCTGAGTAAGTTCCGAGGCGGATACGGAGCAAC 240  
QY 217 TrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAsp 236  
Db 241 TGGATGTGGCGCGTCACACGGGCGACTTCAGCGTATTCGCGTGTATGCCGGTCCGAC 300  
QY 237 AsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAla 256  
Db 301 AACCGCGCGCGGATACAGCAAGGACAAATAAACCTATAAGCCGCTTACTTCGCTGCC 360  
QY 257 ValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySer 276  
Db 361 GTATCCATGCAAGGCTACAGGCTGACGACTATGCCATGACCATCGGTTTCCCGGGCAGT 420  
QY 277 ThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnPro 296  
Db 421 ACGGATCGCTACCTACCTTCTTGGGCTGTGAAGATCGTATCGAAAACGAGACAATCT 480  
QY 297 ArgIleGluValArgGlyIleLysGlnClyIleTrpLysGluAlaMetSerAlaAspGln 316  
Db 481 CGTATCGAAGTTCGGGTATCAAGCAAGGCTCTGGAAGGAGCCATGAGGCGCATCG 540  
QY 317 AlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSer 336  
Db 541 GCTACCGGTATCAATATGCCAGCAAGTATGCTCAGAGTGCTAACTATTTGGAAGAAATTCG 600

QY 337 IleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGlu 356  
Db 601 ATCGGTATGAACCGCGTCTCGCTCGTTCGATAGGTAGTTCGTAAGCGTCCGAGGAA 660  
QY 357 AsqAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrClyAspValLeu 376  
Db 661 AGAGCATTCGACACTGGATCCGTAAAGACGCAAGAGTGTCTCTATGGCATGTATTG 720  
QY 377 SerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeu 396  
Db 721 TCTTCTCTCGAAAGCGCTTATAAGGAAGGAGCAAGGCCAACCGTGAGATGACTATTG 780  
QY 397 SerGluThrLeuPheGlyClyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeu 416  
Db 781 ACCGAGACGCTCTTCGGTGTACCGAGGTGGTTCGTTTTTCACAGTTTTGCCAACCAATG 840  
QY 417 AlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAsp 436  
Db 841 GCTACAAATCCTGATGCTCATGCCGTATCCCAATTCGTTGACGACCAAGTACAAAGAC 900  
QY 437 TyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArg 456  
Db 901 TACCTCCCTCGCTCGACCGTAAGGTGCTGCCCGCCATGCTCGATATTGTACGCCGCGT 960  
QY 457 IleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGly 476  
Db 961 ATCCCTGCCGACAACTCCCGATATATTCAGAATGTATTCGACAGCAAGAAATTCAGAGC 1020  
QY 477 AspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLys 496  
Db 1021 GACACGAGAAGTATGACAGACTTCCTATTCGACAGAGGTGGTTCCTTATAGCCACAG 1080  
QY 497 PheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAsp 516  
Db 1081 TTCCATGCTCATGCTCAAGTCCATGGACAGGAAAGTTTGCAGGCTATCGAGAAAGAT 1140  
QY 517 ProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAla 536  
Db 1141 CCGCGAGTAGAGCTTTCACAGAGCGTAATAGTCTGCTCGCGCTATTTTCAGCGCGATCG 1200  
QY 537 MetaLysAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGlu 556  
Db 1201 ATGGCAATGCTATGCTATGAGAGGCAAGGCTCTTTCTTTTCCGCGTTCGCTGAG 1260  
QY 557 MetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGly 576  
Db 1261 ATGTACCCCGGACGCTCTCTCGGAGCGATGCCAACTTCACCATGCGTATGAGCTACGCG 1320  
QY 577 SerIleLysGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHisThrThrGlyLys 596  
Db 1321 TCCATCAAGGATATGAACCGCAGGCGTCCCTGGTACAACTATCATACGACAGGCAAG 1380  
QY 597 GlyValLeuGluLysGlnAspPro-LysSerAspGluPheAlaValGlnGluAsnIleL 616  
Db 1381 CGCGTATTGGAGAAGCAGGATTCCTTAAGACCGATGAGTTTTCGCTACAGGAGAAATATCC 1440  
QY 616 euAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleA 636  
Db 1441 TCGACCTCTTCGCGACCAAAACATATGCTCGCTATGCGGAGAACGCGTACCATATCG 1500  
QY 636 laPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysA 656  
Db 1501 CTTTCTTATCGAACACACATACGCGCGTAACTCCGCTAGCCCGCTATTCGATTAAGA 1560  
QY 656 snGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleG 676  
Db 1561 ACGGCGCTCTGATCGCTCTGCTTCGATGCAACTCGGAAAGCTATGAGTGGTGCATCG 1620  
QY 676 luPheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetI 696  
Db 1621 AGTTCCGAACCGGATCTCGACGGCAACAATACGCGTGGACATCCGCTACGTTCTCTTCATGA 1680

QY 696 leAspLysTrpGlyGlnCysProArgLeuLeuGlnGluLeuLysLeuIle 712  
 Db 1681 TTGACAAATGGGTGAGTGCCTCCCGCTCATCAAGAGCTGAAGTTGATC 1730

RESULT 2  
 US-09-221-017B-382  
 ; Sequence 382, Application us/09221017B  
 ; Patent No. 6444799  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ross, Bruce C.  
 ; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 1120  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORRISON & FOERSTER  
 ; STREET: 755 PAGE MILL ROAD  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304-1018  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: Windows  
 ; SOFTWARE: FastSeq for Windows Version 2.0b  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/221,017B  
 ; FILING DATE: 23-DEC-1998  
 ; CLASSIFICATION:  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: P1182  
 ; FILING DATE: 31-DEC-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: P1546  
 ; FILING DATE: 30-JAN-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: P2911  
 ; FILING DATE: 09-APR-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/AU98/01023  
 ; FILING DATE: 10-DEC-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Monroy, Gladys H  
 ; REGISTRATION NUMBER: 32,430  
 ; REFERENCE/DOCKET NUMBER: 27340-20021.00  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-813-5600  
 ; TELEFAX: 650-494-0792  
 ; TELEX: 706141  
 ; INFORMATION FOR SEQ ID NO: 382:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1317 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: circular  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: UNKNOWN  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: PORPHYROMONAS GINGIVALIS  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 1...1317  
 ; US-09-221-017B-382

Alignment Scores:  
 Pred. No.: 1.07e-71 Length: 1317  
 Score: 743.50 Matches: 168  
 Percent Similarity: 55.38% Conservative: 74  
 Best Local Similarity: 38.44% Mismatches: 172  
 Query Match: 19.99% Indels: 23  
 Db: 4 Gaps: 7

US-10-008-355-2 (1-712) x US-09-221-017B-382 (1-1317)

QY 41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60  
 Db 16 CAGATCAAGAGCGCGCCTTAAGATGAAGGAATACGACCTTTATATATCCCAACGCACA 75  
 QY 61 SerIleAlaAsnAlaValValIlePheGlyGlyCysThrGlyIleThrValSerAsp 80  
 Db 76 TCGCTGAAAGATGCCGTAGTGTCTTTCGACGAGGAGTACGCGGCGAGGTGCTTCGGAT 135  
 QY 81 GlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100  
 Db 136 CGCGGACTGGTACTGACCAATCCACTCGGATACGATATGATCCAGGCTCACAGCAGC 195  
 QY 101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluPro 120  
 Db 196 CTCGAGCATAACTATCTCGAAATGGATTTCGGCGATGAGAGAAGCGGATGAATTACCG 255  
 QY 121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140  
 Db 256 AACAAGGATATTTCCGTGGTATTTCATCGACAAGATCGAAGATGTCAAGATACGTCAAG 315  
 QY 141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160  
 Db 316 AAGATCTCAAGCCCATCAAGATCCCAAGCATGCGACTACCTCTCTCCGAAGTACCTG 375  
 QY 161 GlnGluLeuAlaLysLysGlu-----AsnAlaAspGluAsnGlnLeuCys 175  
 Db 376 CAAAGTTGGCTGACAAGAAGCGCGCAAGAACTTTCTGCCAAGAATCCGGGCTTTCC 435  
 QY 176 Ile---ValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPhe 194  
 Db 436 GTAGAGATCAAGCCCTTCTATGGGGCAATCTCTACCTGATGTTTACCAAAAGACTTAT 495  
 QY 195 LysAspValArgMetValPheAlaProProSerSerValGlyLysPheGlyAspThr 214  
 Db 496 ACGGATGTTGACTGGTGGGAGCACCTCCACCACGACATTCGCAAAATTCGGTCCGATACG 555  
 QY 215 AspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGly 234  
 Db 556 GACAACCTGGATCTGGCTCGTCACTATGCGGACTTCTCCATCTTCCGTATATATATCGGAC 615  
 QY 235 AlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPhe 254  
 Db 616 AAGATGGCAATCCGACCATCACTCTGAAGATAATGTCGCTCAAGCCCAAGCGTTTC 675  
 QY 255 AlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhePro 274  
 Db 676 TTCAATATCTCCCTGGTGGAGTACAAGAGAACGACTACGCCATGATATAGTGGTTCCCC 735  
 QY 275 GlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsn 294  
 Db 736 GGTACTAGCGCACCGCTATTTCCAGGCTTCCGCAAGTAGACGAATGGAAAGCATCGACAAC 795  
 QY 295 AsnProArgIleGluValArgGlyIleLysGlnGlyIleThrLysGluAlaMetSerAla 314  
 Db 796 GATATTCGCATCCGATCGGTGATATTCGTCAGGGTTCATGCTCAGGAAATGCTGCC 855  
 QY 315 AspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTriLys 334  
 Db 856 GATCTCTAGATCAAAATCATGTATTACGTAATATGCGCTTCGACAAATGCTTACAAA 915  
 QY 335 AsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAla 354  
 Db 916 CGTGCTATAGTGCCCACTGGCGCATCAAGACACGCGCTCGGTCAAAACAAACAGCGC 975  
 QY 355 GluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAsp 374  
 Db 976 ATGAGGACAGGCTGATCGCATGGGAGCGAGCGAGGCT---ACTCCTCGATATGAAGAG 1032  
 QY 375 ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThr 394  
 Db 1033 GCCGTACACGAATCGATGCTACGGTAGCCAGCGTGGGATCTCCGCCGCTGTTATTGG 1092

QY	395	TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArg-----PheAla	410
Db	1093	ATGATAGAAGAGGCATCATCGTGGGATCGAGTTTGCCCTTCCTATCCCAACGAA	1152
QY	411	GlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeu	430
Db	1153	GATGACACGAAAGCTTCGAAGGCAAT---GATGCTTCGGCT-----CGCAAGAGGCG	1203
QY	431	AspAspLysTyr-----LysAspTyrLeuProSerLeu	441
Db	1204	ATCGATAAGATTCGTACACGCTACAGCAAAATTTGCCAACAAAGGACTACAGTGCAGAGGTG	1263
QY	442	AspArgLysValLeuProAlaMetLeuAspIleValArgArgIlePro	458
Db	1264	GACACAGAAATGGCCGTAGCCATGCTGACGGGAAATACCTCAAGGAAATGCC	1314
RESULT 3			
US-09-221-017B-1045			
; Sequence 1045, Application US/09221017B			
; Patent No. 6444759			
; GENERAL INFORMATION:			
; APPLICANT: Ross, Bruce C.			
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF			
; NUMBER OF SEQUENCES: 1120			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: MORRISON & FOERSTER			
; STREET: 755 PAGE MILL ROAD			
; CITY: Palo Alto			
; STATE: CA			
; COUNTRY: USA			
; ZIP: 94304-1018			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette			
; COMPUTER: IBM Compatible			
; OPERATING SYSTEM: Windows			
; SOFTWARE: FastSeq for Windows Version 2.0b			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/09/221.017B			
; FILING DATE: 23-DEC-1998			
; CLASSIFICATION:			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: PP1182			
; FILING DATE: 31-DEC-1997			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: PP1546			
; FILING DATE: 30-JAN-1998			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: PP2911			
; FILING DATE: 09-APR-1998			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: PCT/AU98/01023			
; FILING DATE: 10-DEC-1998			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Monroy, Gladys H			
; REGISTRATION NUMBER: 32,430			
; REFERENCE/DOCKET NUMBER: 27340-20021.00			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 650-813-5600			
; TELEFAX: 650-494-0792			
; TELEX: 706141			
; INFORMATION FOR SEQ ID NO: 1045:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 2394 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: circular			
; MOLECULE TYPE: DNA (genomic)			
; HYPOTHETICAL: NO			
; ANTI-SENSE: UNKNOWN			
; ORIGINAL SOURCE:			
; ORGANISM: PORPHYROMONAS GINGIVALIS			
; FEATURE:			
; NAME/KEY: misc_feature			

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P11182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P11546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 561 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1...561  
US-09-221-017B-12

Alignment Scores:  
Pred. No.: 5.28e-20 Length: 561  
Score: 265.50 Matches: 59  
Percent Similarity: 96.83% Conservative: 2  
Best Local Similarity: 93.65% Mismatches: 2  
Query Match: 7.14% Indels: 1  
DB: 4 Gaps: 0

US-10-008-355-2 (1-712) x US-09-221-017B-12 (1-561)

Qy 1 MetGlnMetLysLeuLysSerIleLeuGlyAlaAlaLeuLeuGlyAlaSerGly 20  
Db 366 ATGCAATGAATTAATAAGATTCTTCGGAGACGCTGCTGTGGGTGAAGG 425  
Qy 21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlnGluAsnLeuAsp 40  
Db 426 GTACCAAGCCGACAAAGGCATGTGCTCTCAACCAACTCAATCAGGAGATCTGGAT 485  
Qy 41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60  
Db 486 CGAATGCTGAGCTCGGCTTTACCT--CCCGTGGATTCTGCTACAGTTTCGACAGCCG 543  
Qy 61 SerIleAla 63  
Db 544 TCCATGCCA 552

RESULT 5

US-08-953-492-1

; Sequence 1, Application US/08953492

Patent No. 5849555  
GENERAL INFORMATION:  
APPLICANT: Brown, James  
APPLICANT: Jaworski, Deborah  
APPLICANT: Lawlor, Elizabeth  
APPLICANT: Wang, Min  
TITLE OF INVENTION: NOVEL vals  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/953,492  
FILING DATE: 17-OCT-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/844,064  
FILING DATE: 18-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9607991.8  
FILING DATE: 18-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31458-4/1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2652 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
US-08-953-492-1

Alignment Scores:  
Pred. No.: 0.000381 Length: 2652  
Score: 130.00 Matches: 106  
Percent Similarity: 32.81% Conservative: 62  
Best Local Similarity: 20.70% Mismatches: 166  
Query Match: 3.50% Indels: 178  
DB: 2 Gaps: 25

US-10-008-355-2 (1-712) x US-08-953-492-1 (1-2652)

Qy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203  
Db 76 GATGTTTTCAGCCCTTCAGCGGATCAAAAGCGTAAAGCCTTATTCAATCGTTATTCACCA 135  
Qy 204 ProSerSerValGlyLysPheGlyAspThrAspAsnTrpMetTrpProArgHisThr 223  
Db 136 CCAACCGTTACAGGTAAACTT-----CACCTT 162  
Qy 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlyTyrSer 243  
Db 163 GGT-----CAGCT 171  
Qy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263  
Db 172 TGGGATACA---ACTTTGCAAGATATTATCATCCGTCAAAACGATGCAAGGTTTTGAT 228

Qy 264 AlaAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282  
Db 229 -----ACCTTTGGCTTCTGGATGGACCGACGAGGATGCCACT 270  
Qy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302  
Db 271 CAGGCTAAGGTAGAGGACGCTCGCTGGTGGAGGCAATTCCTCGTAGACTATTCCTCGTGAGCGTTTC 330  
Qy 303 -----IleGluGlnGlyIleTrpLysAlaMetSerAlaAspGlnAlaThr 318  
Db 331 GAGCTCTTTCTGACCAAGACTCTGGGAATGGAAAGAC-----GAATATGCCACT 378  
Qy 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337  
Db 379 ACTATCAAGGAACAATGGGCAAGATGGGCTCTCTGTAGACTATTCCTCGTGAGCGTTTC 438  
Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357  
Db 439 ACTCTTGACGAAGGTTTCTCA-----AAAGCTGTTCTGTAAG 474  
Qy 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373  
Db 475 GTCCTTTGGACCTTTACAAAGAAAGCTGGATCTACCGTGGTGAGTTTATCATCAACTGG 534  
Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLys 388  
Db 535 GACCCAGCAGCTCGCACAGCCCTTCTCATATTGAGGTGATTCACAGGATGTGAAGGT 594  
Qy 389 AlaAsnArgGluMetThrTyrLeu-----GACCCGCGCTAC 684  
Db 595 GCCTTCTACCACTGAATATACATGCTGGAAGATGTTCCACGCGTCTTGAAGTTGCTACA 654  
Qy 397 -----SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414  
Db 655 ACTCGCTCTGAGACCATGTTTGGGACGTT----- 684  
Qy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434  
Db 685 GCGGTGGGGTCAACCCAGAA-----GACCCGCGCTAC 717  
Qy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448  
Db 718 AAGGACTTGTGTTAAATAATGTATCTCTTCCA---ATCGCTAATAAACTCATCCCAATC 774  
Qy 449 MetLeuAsp-----IleValArgArg 455  
Db 775 GTTGAGATGAGCAGCAGATCCTGAGTTTGGTACTGTGTCGTGAAATCACACCTGCC 834  
Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475  
Db 835 CACGATCCCAATGACTTCTTGGTGGCCACGTCATACTTCCCAAGTCAACGTCATG 894  
Qy 476 GlyAspThrLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495  
Db 895 AACGACGACGAACCATCAATGACTTGGCCCTTGAATTTTCAGGCATG-----GAC 945  
Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515  
Db 946 CGTTTGAAGCTCGT-----AAGCAGCTCGTGTGCTAAGTTGGAAGAAATC 990  
Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaArgAla----- 531  
Db 991 GGTGCGCTCGTCAAAATCGAAAACGTTCTCCACAGTGTGTCACGAGCGTACAGGT 1050  
Qy 532 -----IleGlnAla 534  
Db 1051 GTTGTGGTGAACCTCGCTTCTACTCAATGTTGTCGTAAGATGGACCAATGCTGAAG 1110  
Qy 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554  
Db 1111 AACGCCATTGCCAACCAAGACACAGAGGACAAAG----- 1143  
Qy 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565

Db 1144 GTCGAATTTCTACCCACCTCGTTTCAACGATACCTTCTCAATGGATGGAATAATGTCAC 1203  
Qy 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585  
Db 1204 GACTGGGTATCTTCGTGTCAGCTCTGGTGGGTAC-----CAAATC 1245  
Qy 586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605  
Db 1246 CCTCGCTGGTACAAT-----GCTGATGGTGAATATGTATGTCGGCGAAGAGCTCCAGAA 1299  
Qy 606 SerAspGluPheAlaValGlnGlnGluAsnIleLeuAsp 617  
Db 1300 GGTGACGATGAGCTCAGGACGACGAAGACGCTCTGGAT 1335  
RESULT 6  
US-08-961-527-27/c  
; Sequence 27, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961.527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6004 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-27  
Alignment Scores:  
Pred. No.: 0.00173 Length: 6004  
Score: 129.00 Matches: 116  
Percent Similarity: 33.62% Conservative: 77  
Best Local Similarity: 20.21% Mismatches: 194  
Query Match: 3.47% Indels: 187  
DB: 4 Gaps: 26  
US-10-008-355-2 (1-712) x US-08-961-527-27 (1-6004)  
Qy 131 LysIleValLysValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMet 150  
Db 4315 AAATAATAATACGAATTTGTCTGCAGCAAGCTGCTAAGCAATTCAAATAATAATA 4256  
Qy 151 GluArgLeu-ArgLysAlaGlnGluCysGlnGluAlaLysLysGluAsnAlaAs 170  
Db 4315 AAATAATAATACGAATTTGTCTGCAGCAAGCTGCTAAGCAATTCAAATAATAATA 4256

Db 4255 AACAAATAGAGAAAGGACACACATCTCTAAAGAACTTTCACTAAATACATCCAGC 4196  
Qy 170 pGluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTyrPheLeuIleVa 190  
Db 4195 CGAGGTGAGCGCTGCTGTACCAAAATGG-----CTTGA 4160  
Qy 190 lTyAspValPhe-----LysAspValArgMetValPheAl 202  
Db 4159 TGCAGATGTTTTCAGGCTTCAGCGCATCAAAAGCCTAAGCGCTATTCAATCGTGATCC 4100  
Qy 202 aProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHi 222  
Db 4099 ACCACCAACCTAAGTGGAACTT-----CA 4073  
Qy 222 sThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTy 242  
Db 4072 CCTTGGT-----CA 4064  
Qy 242 rSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTy 262  
Db 4063 CGCTTGGGATACG---ACTTTGCAGGATATCATCATCCGTCAAAAGCGCATGCAAGGCTT 4007  
Qy 262 rLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu-- 281  
Db 4006 TGAT-----ACCTTTGGCTTCCAGGTATGACCAACCGCTGTATTTC 3965  
Qy 282 -ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValAr 301  
Db 3964 GACTCAGGCTAAGCTTGAGGAGCGCTTACGTGGTGAGGGTATTACGCGTWTATGACCTAGG 3905  
Qy 301 gGly-----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAl 317  
Db 3904 TCGTGAGTCTTCTTGACGAAGCTGCGAATGGAAGAC-----GAATATGC 3857  
Qy 317 aThrArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSe 336  
Db 3856 CACTTACATCAAGGAACAATGGGCGACAGATGGGCTTCTGTAGACTATTCTCGTGAGCG 3797  
Qy 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGl 356  
Db 3796 TTTTCACCTCTGATCAAGTTTGTCA-----AAAGCTGTTCG 3761  
Qy 356 uArgAlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTy 372  
Db 3760 TAAGTCTCTTTGTAACCTTTACAAGAAAGGCTGGATCTATCGTGTGATGTTATCATCAA 3701  
Qy 372 rGlyAsp-----ValLeuSerSerLeuCluLysAlaTyrLys----- 384  
Db 3700 CTGGACCCAGCAGCTCGCACAGCCCTTCTGTATGATGATGATTCACAAGGATGTAGA 3641  
Qy 385 -----GluGlyAlaLysAlaAsnArgGluMe 393  
Db 3640 AGGTGCCTTCTACCACATGATATCATCTGGAAGATGGTTACGCGCCCTTGAAGTTGC 3581  
Qy 393 tThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAl 413  
Db 3580 TACAACCTCGTCTGAGACTATGTTGGGAGCTT----- 3547  
Qy 413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLy 433  
Db 3546 -----GCGGTTGCGGTTAATCCAGAA-----GACCGCGC 3518  
Qy 433 sTyTyrLysAsp-----TyrLeuProSerLeuAspArgLysValLeuPr 447  
Db 3517 CTACAAGGACTTGATTGGTAAACACGTCATCTTCCA---ATCCCTAATAACATCATCCC 3461  
Qy 447 oAlaMetLeuAsp-----IleValAr 454  
Db 3460 AATCGTTGGAGATGAGCAGCAGATCCTGAGCTTGGTACTGCTGCGTGAATAATCACACC 3401  
Qy 454 gArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPh 474  
Db 3400 TGCCACGATCCAAATGACTTCTTGGTTGGCCAAACGTCATCAATGTCGCAACAAGTCAACGT 3341

Qy 474 eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe 494  
Db 3340 CATGAACGACGACGCGAAGTGAATGAGCTTGCCTTTTCAATTTTCAAGCATG----- 3289  
Qy 494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGl 514  
Db 3288 -GATCGTGTGTTGAAGCTCGT-----AAGCAGCTCGTGTGCTAAGTTTGAAGA 3245  
Qy 514 uLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531  
Db 3244 AATCGGTGCCCTTGTCAAAATCGAAACACGTGTCCACAGCTTGTGCTCACTCAGAACGTAC 3185  
Qy 532 -----IleGl 533  
Db 3184 AGGTGTCGTAGTTGAGCCACGCTTGTCTACTCAATGTTCTCAAGATGGACCAATTGGC 3125  
Qy 533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGl 553  
Db 3124 TAAGAAGCCCATTCCTTAACCAAGACACAGAGACAAG----- 3088  
Qy 553 yLeuArgGluMetTyrProGlyArgAlaLeuProSer----- 565  
Db 3087 ---GTCAATCTACCCACCTCTTTCACAGATACCTTCTCCTCAATGGATGGAATGT 3032  
Qy 566 ---AspAlaAsnPheThrMetArgMetSerTyrIleLysGlyTyrGluProGl 584  
Db 3031 CCACGACTGGGTTATCTCTCGTCAGCTTGTGGGGTCAC-----CA 2990  
Qy 584 nAspGlyAlaTrpTyrAsnTyrHisThrGlyLysGlyValLeuGluLysGlnAspPr 604  
Db 2989 AATCCCTGCCTGGTACAAT-----GCTGATGTTGAAATGATGTCGGCGAAGAAGCTCC 2936  
Qy 604 oLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 617  
Db 2935 AGAAGGTGACGGATGGACTCAGGACGAAGACGTCTTGGAT 2896

RESULT 7  
US-09-059-584-52  
: Sequence 52, Application US/09059584  
: Patent No. 6440701  
: GENERAL INFORMATION:  
: APPLICANT: Myers, Lisa E  
: APPLICANT: Schryvers, Anthony B  
: APPLICANT: Harkness, Robin E  
: APPLICANT: Loosmore, Sheena M.  
: APPLICANT: Du, Run-Pan  
: APPLICANT: Yang, Yan-Ping  
: APPLICANT: Klein, Michel H  
: TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
: NUMBER OF SEQUENCES: 60  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Sim & McBurney  
: STREET: 6th Floor, 330 University Avenue  
: CITY: Toronto  
: STATE: Ontario  
: COUNTRY: Canada  
: ZIP: M5G 1R7  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/09/059,584  
: FILING DATE: 14-APR-1998  
: CLASSIFICATION:  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/778,570  
: FILING DATE: 03-JAN-1997  
: CLASSIFICATION:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Stewart, Michael I



---

---



Qy 673 GlyAspIleGluPheGluProAsp 680  
 Db 2065 GGGTCATTACACAGATACCGAT 2088

RESULT 8  
 US-08-793-824-1  
 ; Sequence 1, Application US/08793824  
 ; Patent No. 5981838  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Simpson, Christine Lynn  
 ; APPLICANT: Giffard, Philip Morrison  
 ; APPLICANT: Jacques, Nicholas Anthony  
 ; TITLE OF INVENTION: Genetic Manipulation of Plants to  
 ; TITLE OF INVENTION: Increase Stored Carbohydrates  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Griffith Hack & Co  
 ; STREET: Level 8, 168 Walker Street  
 ; CITY: No. 5981838th Sydney  
 ; STATE: New South Wales  
 ; COUNTRY: Australia  
 ; ZIP: 2060  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA: US/08793,824  
 ; FILING DATE:  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: AU PM7643  
 ; FILING DATE: 24-AUG-1994  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 61 2 9957 5944  
 ; TELEFAX: 61 2 957 6288  
 ; TELEX: 26547  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4853 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Streptococcus salivarius  
 ; US-08-793-824-1

Alignment Scores:  
 Pred. No.: 0.0423 Length: 4853  
 Score: 115.00 Matches: 161  
 Percent Similarity: 34.10% Conservative: 106  
 Best Local Similarity: 20.56% Mismatches: 262  
 Query Match: 3.09% Indels: 257  
 DB: 2 Gaps: 40

US-10-008-355-2 (1-712) x US-08-793-824-1 (1-4853)

Qy 70 GlyGlyGlyCysThrGlyIleThrValSerAsp-----GlnGlyLeuIle 84  
 Db 1776 GGTGCACATGGGATGGGATCTCTTCGAGATTGCTTGTGATTTACTTCAAGGAAAAATA 1835

Qy 85 PheThrAsnHis-----HisCysGlyTyr-GlyAlaIleGlnSerGlnSe 99  
 Db 1836 CGGTGCGCAGATAATGAAGCAATGCCATGCTCATTTGTCAATCCTTGAAGCTTGCTC 1895

Qy 99 rThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluGluLe 119  
 Db 1896 ATACAACGACCACCAATACAAACAGGAT-----ACCAAGGTTGCTCAGTT 1940

Qy 119 uProIleProGly-----LeuSerValIlyTyrLeuArgLys----- 131  
 Db 1941 GTCATCGACAATCCACTACTGCTGAAACGCTTTTGACGACTTTCTTCGTAAGACAAATA 2000

Qy 131 ----- 131

Db 2001 CCGTGGCAGCTTGGAGCGGTGATTACTACTCTTAAACAATCGTTCAAGTGAGCAGAA 2060

Qy 132 -----IleValIlyValThrAsp----- 137  
 Db 2061 ACACAGCCACGTGATGCCAATTATATTTCTGCGGAGCCCATGATAGTGAAGTCAAGC 2120

Qy 138 -----LysValGluGlyGlnLeuLysGlyIleThrAspG1 149  
 Db 2121 TGTTTTGGCTAATATCATCAGCAGCATTAATCCAAAACAGATGCTTTCACCTTCAC 2180

Qy 149 uMetGluArgLeuArgLysAlaGlnGluValCysGln---GluLeuAlaLys-----Ly 166  
 Db 2181 TATGATGAGCTCAACAGGCTTCGAAATCTACAATCGGACATCGCGAAGGCTGATAA 2240

Qy 166 sGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSerAsnAsnGluTy 186  
 Db 2241 AAAGTACACCCAGTACAATATCCAGCTGCCTATGCCACAATGTTGACCAACAGGATAG 2300

Qy 186 rPheLeuIleValTyr-----AspValPheLysAspValArgMetValPheAlaProPr 204  
 Db 2301 TATCACTCGCTTTTACTACGGGACCTCTTTACCGACCATGCCCCAATACATCGCTGA 2360

Qy 204 oSer----- 205  
 Db 2361 ATCACCGTACTATAATGCCATCGATGCCCTGCTCGCTGCATCAATATGTAGCAGG 2420

Qy 206 -----SerVa 207

Db 2421 TGGTCAAGACATGAAGTTACTAAGCTTAATGGCTATGAAATCATGTCATCTGTCGGT 2480

Qy 207 lGlyLysPheGlyGlyAspThrAspAsnTrpMetTrp-----ProAr 221  
 Db 2481 TGGTAAG---GGGCGAAGAAAGCAACCA-CCTTGGTACTCTGCTGAACACGCAACCA 2536

Qy 221 gHisThrGlyAspPheSer---ValPheArg-ValTyrAlaGlyAlaAspAsnArg---- 238  
 Db 2537 GAATGCTGCTCTTACAGCCAAACCGTCCAGATATGAAGTTGGGAGTAATGATCATCTCG 2596

Qy 239 -----ProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAla- 256  
 Db 2597 TAGTCAATATGGGAGCTGCCCAAAAATCAGGCTTACCGTCCATTGCTTCTCAGCAAA 2656

Qy 257 -----V 257

Db 2657 CGACAGGCTTTCGACCTACCTCAAGGATTCTGTATGTACCAGCTGGATTGGTTCGTATA 2716

Qy 257 alSerMetGlnGly-----TyrLysAlaAspAspTyrAla-----MetThrIleG 272  
 Db 2717 CGGACAATCAAGGAACTTGACCTTTCAGCGCAGATGATATGCTGCTCATTTCAACCGTT 2776

Qy 272 lPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluA 292  
 Db 2777 AAGTTTTCAGGT-----TATTTGGCAGTTGGGTGCCAGTGGGTGCTCAGAAA 2824

Qy 292 snGluAsnAspProArgIleGluValArgGlyIleLysGlnGly-----IleTgPLysG 310  
 Db 2825 ACCAA---GATGCCCGAACCAAGGCTTCGAGCACCAGAAAGGTCGAGGCTCTTTGAAT 2881

Qy 310 luAlaMetSerAlaAspGlnAlaThr-----ArgI 320  
 Db 2882 CATCAGCAGCTCTTCATTACACAAGTCATTACGAAGGCTTCTCAAACTTCCAAAGATT 2941

Qy 320 leLysTyrAlaSerLysTyr-----AlaGlnSerAlaAsnTyrTrpLysA 335  
 Db 2942 TTAAGACACCAAGTCAGTACACCAACCGCTGTTATTGCTCAAAATGCCAACTCTTCAA 3000

QY 335 snSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg---LysArgA 354  
Db 3001 -----GAGTGGGAATCACTTCTTCAATTTGGCCACAGATGTGTCTA 3046  
QY 354 laGlulArgAlaPheAlaAspTrpIleArgLysnGly----- 367  
Db 3047 GCAAGACGGCACATTTCTTGGATTCTATCATGAAATGGCTACGCTTCGAGGACCGCT 3106  
QY 368 -----LysSerAlaValTyrGly-----AspValLeuSerSerL 379  
Db 3107 ACGATATTGCCATGACGACAGACAACAAATACGTTCACTCAAGACCTCATGGATGCC 3166  
QY 379 euGlulYalaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluT 399  
Db 3167 TTCGTGCCCTTCACGGAAGAGGTATCTCAGCCATCTCGAC-----TGGGTTCCAGATC 3220  
QY 399 hrLeuPheGly-----GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuA 417  
Db 3221 AAATTTACAATCTCCTCGAAAGAGAGTGTGAACA---GCTTCTGTACCAATAGCTAG 3277  
QY 417 laThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeu---AspAspLysTyrLysA 436  
Db 3278 GTACA---CCAGCTCCAATGTGAATCTACAATAGCTCTACGACGCTAAACACGTA 3334  
QY 436 spTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgA 456  
Db 3335 CCTTTGGAATGACTTCCAAGGACAGTACGGTGGTCTCTTGTATGAATTTGAAGGCAA 3394  
QY 456 rgIleProAlaAspLysLeuProAspIlePheLysnValIleAspLysLysPheLysG 476  
Db 3395 AATACCCAGCA-----ATCTTGAGCGGTGCA-GAT----- 3425  
QY 476 lYAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspL 496  
Db 3426 -----TTCAACGGCGGCTAATTTGCTACCAATGAG 3456  
QY 496 yPheHisAlaMetLeuLys---SerMetAspLysGluLysPheAlaLysAlaIleGluL 515  
Db 3457 AAAATCAGCAATGCTGAGCAAGTATTTCAATGGAAGCAATATCCAGGTACTGGAG-- 3514  
QY 515 sAspProAlaValGluLeuSerLysSerValIleAlaAlaArgAlaIleGlnAlaAs 535  
Db 3515 -----CTCGCTAT-GTCTCAACAGATAA 3536  
QY 535 pAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeuAr 555  
Db 3537 CGCTACCAACCAATACTTCAAGCGTCAACAGCGTCAACCTTCTCTCAACAAATGAC 3596  
QY 555 gLuuMetTyrProGly-----ArgAlaLeuProSerAspAlaAsnPheThrMetArgMe 573  
Db 3597 TGAATTTACTGGAAGTGGTTTCCGTAGGTTGGAGATGAT-----GT 3638  
QY 573 tSerTyrGlySerIleLysGlyTyrGluPro-----GlnAspGlyAla-- 587  
Db 3639 CCAATACCTCTCAATTTGGTGGTCTGCTGAAGAAATACCTTTATCAAGTGGTGCCAA 3698  
QY 588 ----TrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSe 606  
Db 3699 CCAGTGGTATTACTTTGATAAGAAATGGCAACATGCTCACAGGTGACAGGTCAATGATCG 3758  
QY 606 rAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyAr 626  
Db 3759 CAAGAATACTTCTTCTTAGACAATGGTCTCCAGCTA-----CGTCAATGCTCTTCG 3809  
QY 626 gTyrAlaGluAsnGlyGlnLeuHisIle-----AlaPheLeuSe 639  
Db 3810 CCAAGGTAGTGTATGTCATGTGTATTATTACATCCTTAAGGGGTTACGGCTTTAAACGG 3869  
QY 639 rAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyArgLe 659  
Db 3870 ATTTTATGATTTTGGCGGTCTCTCGCAAGACGCTTCGCTACITTTGATGGCAAGGTCAAT 3929  
QY 659 u 659

Db 3930 G 3930

## RESULT 9

US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

## Alignment Scores:

Pred. No.: 4,32e+03 Length: 4411529  
Score: 111.00 Matches: 137  
Percent Similarity: 29.45% Conservative: 60  
Best Local Similarity: 20.48% Mismatches: 187  
Query Match: 2.98% Indels: 286  
Gaps: 35

US-10-008-355-2 (1-712) x US-09-103-840A-1 (1-4411529)

QY 75 GlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHisHisCysGlyTyrGlyAla 94  
Db 4023701 GGTCCACGGTG-----CAGGGCTA-----CCGCATATCGGCAT----- 4023666  
QY 95 IleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThr 114  
Db 4023665 -----GTCCGACAGGGGTGCTTGCATCTCTGCGCCGATGGTTCCTCGCGCT--- 4023615  
QY 115 MetGlyGluGluLeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLys 134  
Db 4023614 -----GGCTACGACGTGCGGTTTATCCGCAACGTGACCGAC 4023579  
QY 135 ValThrAspLysValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArg 154  
Db 4023578 ATCGAAGACAAAGATC----- 4023564  
QY 155 LysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeu 174  
Db 4023563 -----CTGCCAAGCCCGCGCGCGGC----- 4023540  
QY 175 CysIleValGluProPheTyr-----SerAsnAsnGluTyrPheLeuIleVal 190  
Db 4023539 -----CGGCCGTGGTGGAGTGGCGGCTACCCACGAGCGTGCCTTACCGCGGCC 4023489  
QY 191 TyrAspValPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPhe 210  
Db 4023488 TACGACGCTCTG---GACGTC-----TTGCCGCGCTCCGCGGAG----- 4023453  
QY 211 GlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe--- 229  
Db 4023452 -----CCGCGCCACCGCGGCATATCACCAGATGATC 4023420  
QY 230 -----ArgValTyrAlaGlyAlaAspAsnArgPro 239  
Db 4023419 GAGATGATCGAGCGCTGATCCAGCGCGCGCATACCGGTGGCGGTGAC----- 4023366  
QY 240 AlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaValSer--- 258

Db 4023365 -----|||||  
Qy 259 -----MetGlnGlyTyrLysAlaAspTyrAlaMetThrIleGlyPhe 273  
Db 4023341 CCGAGTACGGCCAGTTATCCGTCACAGATGATGATCCATCAGCGCGAAGGTGTG 4023282  
Qy 274 ---ProGlySerThrAsp---ArgTyrLeuThrSerTrpGlyValGlu-AspArgIleG1 291  
Db 4023281 GTCGCCGGAAGCGGACGACGCGACTTCACTTTGTGGAAGCGGAAAAGCGGGTG-- 4023224  
Qy 291 uAsnGluAsnProArgIleGluValArgGlyIleLysGlnGlyLeuPheLysGluAl 311  
Db 4023223 -AACCGTCGTGGCCGACGCGTGGGCGCGCGTCCGGGTGGCATCTGG----- 4023173  
Qy 311 aMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAs 331  
Db 4023172 -----|||||  
Qy 331 nTyrTrp-----LysAsnSerIleGlyMetAsnArgGlyLeuAlaAr 345  
Db 4023164 GCAATGGCTCGCAGCTATCTCGGGCGGAATTCGATATCCATTCG-CGGTGAATGGATT 4023106  
Qy 345 gLeu-----AspValIleGlyArgLysArgAlaGluGluArgAlaPheAl 360  
Db 4023105 AGTCTTCCCGCATCATGAGAACGAAATGCCAGCGCGCGCGGTGATGGGTTCGC 4023046  
Qy 360 a-----AspTrpIleArgLysAsnGlyLys-----SerAlaValTy 372  
Db 4023045 CCGCTACTGGCTGCACAAATGGTGGTGGTGGCGGGGAGAAATGAGCAATGCGCT 4022986  
Qy 372 rGlyAspValLeuSer-----SerLeuGluLysAlaTy 383  
Db 4022985 GGGCAACGTGTTCTCCATCGCGCGATGTTGACGCGGTGGCGCGCTGAGTGGCTTA 4022926  
Qy 383 rLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyG1 403  
Db 4022925 TTACCTGGGAGCGCGCACTACCGGTCGATGCTCGAGTCTCCGAGACC----- 4022877  
Qy 403 yThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHi 423  
Db 4022877 -----|||||  
Qy 423 sAlaGlyLeuLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspAr 443  
Db 4022876 -----GCTATGACGAGTGGGTGAAGCCTATGTCGG----- 4022844  
Qy 443 gLysValLeuProAlaMetLeuAspIleValArgArgIle-----ProAl 459  
Db 4022843 -----CTAGAGGACTTCTGCACCGCTTCGACCGGGTGGCGCGCTGTGCCCGCG 4022791  
Qy 459 aAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLy 479  
Db 4022790 CGATCCGACCCCA----- 4022778  
Qy 479 sLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAl 499  
Db 4022778 -----|||||  
Qy 499 aMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaVa 519  
Db 4022777 -----CGGTTCGCGGAGCGCTCGACGACGACCTGTCCGT 4022743  
Qy 519 lGluLeuSerLysSerValIleAlaAlaAraArgAlaIleGlnAlaAspAlaMetAlaAs 539  
Db 4022742 TCCGATCGCGCTCGCGAGATTCACCACTGCGGGG----- 4022706  
Qy 539 nAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPr 559  
Db 4022705 -----GAAGGCAACCGGCGACTCGATGCGCGGCGACCGAC----- 4022670  
Qy 559 oGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLy 579  
||||| ||| |||

Db 4022669 ----GGGCGCTTCGGAAGCGCTAGCCGATTCGGCGGATGATG-----GGCATCT 4022623  
Qy 579 sGlyTyrGluProGlnAspGlyAlaTyrAsnTyrHisThrThrGlyLysGlyValLe 599  
Db 4022622 GGGCTGTGACCGCTCGACCGCGTGG----- 4022595  
Qy 599 uGluLysAlaAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu-- 618  
Db 4022594 -GAATCCCGAGAGAAACCTCGGCGCGCTGCGTGTCTGTCTCCAGGCTGA 4022536  
Qy 619 -PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyClnLeuHisIleAlaPheLe 638  
Db 4022535 ACTACAGAATCGGGAAAGCGCGCGAGCGCAACTGGGCGCTC----- 4022490  
Qy 638 userAsnAsnAspIleThrGly-----GlyAsnSerGlySerProValPheAspLysAs 656  
Db 4022489 ----GCCGACGAGATCCGGGTGGCTGAAGAGCGCGCATCGAGTCCACCGAC----- 4022439  
Qy 656 nGlyArgLeuIleGlyLeuAlaPheAspGly---AsnTrpGluAlaMetSerGlyAspI1 675  
Db 4022438 -----ACCGCCGACGCGCCACAGTGTGCTGCTGCTGGTGTGCACAC 4022398  
Qy 675 eGluPheGluProAspLeuGlnArg 683  
Db 4022397 CAAGTAGATGCCCGTAACCTCTCGG 4022373  
RESULT 10  
US-08-961-527-134  
Sequence 134, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12665 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-134  
Alignment Scores:  
Pred. No.: 0.578 Length: 12665  
Score: 110.50 Matches: 132  
Percent Similarity: 35.47% Conservative: 100  
Best Local Similarity: 20.18% Mismatches: 246

Query Match:	2.978	Indels: .	176
DB:	4	Gaps:	30

US-10-008-355-2 (1-712) x US-08-961-527-134 (1-12665)

Qy	127	Lys	Tyr	Leu	Arg	Lys	Ile	Val	-----Lys	Val	Thr	Asp	Lys	Val	Ile	Glu	Gly	141	
Db	808	GA	TAT	GT	TAAAAA	TA	GT	GGT	GAG	AGCT	AT	GCA	AA	TCA	AACT	TAAAA	AGC	AT	867
Qy	142	Gln	Leu	Lys	Gly	Ile	Thr	Asp	Glu	Met	Glu	Arg	Leu	Arg	Lys	Ala	Gln	Ile	161
Db	868	ATT	ACT	GT	TAG	CTCT	TAC	CGT	TGA	CGT	TGA	CA	CA	AT	TAA	G-----	ACG	GAT	921
Qy	162	Glu	Leu	Ala	Lys	Lys	Glu	Asn	Ala	Asp	Glu	Asn	Gln	Leu	Cys	Ile	Val	Glu	181
Db	922	AAA	TAG	TTC	GAAT	CA	ACCT	CA	GAA	AGC	CA	CTAC	GAT	ACT	GAT	GAT	GAG	AT	981
Qy	182	Ser	Asn	Asn	Glu	Tyr	Thr	Leu	Ile	Val	Tyr	Asp	Val	Leu	Asp	Val	Arg	Met	201
Db	982	AAA	GT	AG	TG	AT	GAA	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	993
Qy	202	Ala	Pro	Ser	Ser	Val	Gly	Lys	Phe	Gly	Lys	Asp	Thr	Asp	Asn	Trp	Met	Trp	221
Db	994	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1029
Qy	222	His	Thr	Gly	Asp	Phe	Ser	Val	Phe	Arg	Val	Tyr	Ala	Gly	Ala	Asp	Asn	Arg	239
Db	1030	TCA	AGT	TC	AG	ACT	CT	TCC	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1062
Qy	240	Ala	Glu	Tyr	Ser	Lys	Asp	Asn	Lys	Pro	Tyr	Lys	Pro	-----	-----	-----	-----	-----	251
Db	1063	TC	AG	TAC	AG	CG	AG	CA	AG	CA	AG	CG	CA	AG	CA	AG	CA	AG	1122
Qy	252	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	269
Db	1123	AAG	AGG	TTC	AAG	AACT	CG	AG	CA	AG	CA	AG	CA	AG	CA	AG	CA	AG	1179
Qy	270	Thr	Ile	Gly	Phe	Pro	Gly	Ser	Thr	Asp	Arg	Tyr	Leu	Thr	Ser	Trp	Gly	Val	289
Db	1180	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	1233
Qy	290	Ile	Gly	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	298
Db	1234	GT	GAG	AGT	TAAAA	AA	CGG	AGCT	GT	CACT	AGT	TAA	AA	GT	GAA	AGCT	CA	AA	1293
Qy	299	Glu	Val	Arg	Gly	Ile	Lys	Gln	Gly	Ile	Trp	Lys	Glu	Ala	Met	Ser	Ala</		

Qy	423	HisAlaGlyIleLeuIysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAsp	442
Db	1702	AAAAACGGAGCTTGAACCTAGTAAAGAGGAGGAGTAAAGAA-----CCTCGAAACG	1755
Qy	443	ArgLysValLeuProIleMetLeuAspIleValArgArgArgIleProAlaAspLysLeu	462
Db	1756	GAAAGAGTTAAGCAGCAAAAGCGGAAGTTGAGAGTTAAAGAAAGCTGAGGCTACAAGGTTA	1815
Qy	463	ProAspIlePheLysAsnValIleAspLysPheLysPheLysGlyAspThrLysTyrAla	482
Db	1816	-----GAAANAAATCAAGACAGATCTGTAANAACGACAA 1848	
Qy	483	AspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys	502
Db	1849	GAA-----GAAGCTAAACGAAACGACGACGACGACGAA 1878	
Qy	503	SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer	522
Db	1879	GATAAGCTTAAGAGAAACAGCTGAACACACACACGCGCGCGCTCCAAAACGACAA 1938	
Qy	523	LysSerValIleAla-----AlaAlaArgAlaIleGlnAlaAspAlaMet	537
Db	1939	AAACCGCTCCAGCTCCAAACACGAGATCCAGCTGAACACCAACGACGACGAAACCA 1998	
Qy	538	AlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhe-----PheAlaGlyLeu	554
Db	1999	GCTGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAATCGCTTG 2058	
Qy	555	ArgGluMetTyrProGlyArg-----AlaLeuProSerAspAlaAsnPheThr	570
Db	2059	ACTCAACGACCGCCAAACAACTGAAACACGACCAACACCATCTACTCCAAA----- 2112	
Qy	571	MetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTyrPyrAsn	590
Db	2113	-----ACAGCTGGAAACAGAAACGGTATGTGGTACTTC 2148	
Qy	591	TyrHisThrThrGlyLys-----GlyValLeuGluLysGlnAspProLysSerAsp	607
Db	2149	TACAATACTGATGTTCAATGGCGCAGGATGGCTCCA----- 2187	
Qy	608	GluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrClyArgTyr	627
Db	2188	-----AACAAATGGCTCATGG 2202	
Qy	628	-----AlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsp-----	642
Db	2203	TACTACCTCAACAGCAATGGCGTATGGCGACAGGATGGCTCCAAACCAATGGTTCATGG 2262	
Qy	643	-----IleThrGly-----GlyAsnSerGlySerPro	651
Db	2263	TACTATTCAACGCTAATGTTCAATGGCAACAGGATGGCTCCAAACCAATGGTTCATGG 2322	
Qy	652	ValPhe---AspLysAsnGlyArgLeuIle-----GlyLeuAlaPheAspClyAsnTrp	668
Db	2323	TACTACCTAAACGCTAATGGTTCATATGGCGACAGGATGGCTCCAATACAATGGCTCATGG 2382	
Qy	669	GluAlaMetSerGlyAsp-IleGluPheGluProAspLeuGlnArgThrIleSerValAs	688
Db	2383	TACTACCTAAACGCTAATGGTTCATATGGCGACAGGATGGCTCCAATACAATGGCTCATGG 2442	
Qy	688	pIleArgTyrValLeuPheMetIleAspLysTrpGlyGln	701
Db	2443	TACTACCTAAACGCTAATGGTTCATATGGCGACAGGATGGCTCCAATACAATGGCTCATGG 2476	

RESULT 11  
US-09-134-078-14  
: Sequence 14, Application US/09134078  
: Patent No. 6368844  
: GENERAL INFORMATION:  
: APPLICANT: Bylina, Edward J.  
: TITLE OF INVENTION: GLYCOSIDASE ENZYMES  
: NUMBER OF SEQUENCES: 72  
: CORRESPONDENCE ADDRESS:

182	rAsnAsnGluTyrPheLeulleValTyrAspValPheLysAspValArgMetValPheAl	207
Qy	:::	
Db	CGATGGAGAG---CTCGGAGCCGTATATTCTCCAGAGAAGACCATATTTCAGAGCTTGCGT	450
Qy	:	
Db	apropSerSer-----ValGlyLysPheGlyGlyAspThrAsp-----	202
Qy	:	
Db	CCCCGTTCTAAGTGGTAAGAGGTCTCTCTTCAMAAACGGAGAGACACACAGAACCGTA	507
Qy	-----ValGlyLysPheProArgHisThr-----GlyAs	216
Qy	-----ValGlyLysPheProArgHisThr-----GlyAs	216
Db	CCAGGTTGTGAACATGAATACAAGGCAACCGGGTCTCGGAAGCGGTCTTCAAGCGCA	567
Qy	pHe---SerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLy	225
Qy	:	
Db	TCTCAGCGGAGTGTCTTACCTCTAT-----CAGCTCGAAAACCTACCGAA	627
Qy	sAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAl	244
Qy	:	
Db	GATCAGAACAAACCCTCGATCTCT--TATTGGAAGCGGTT-----TACCGCAA	672
Qy	aAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuthrSerTr	264
Qy	:    :::	
Db	CAACCAAGAGAGCCCGTGTGTGATCTTCGCCAGCACAAACCCAGAA-----GGATG	717
Qy	pGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleValArgGlyIleLy	284
Qy	:	
Db	-----GAAAAACGACGAGGGACCGAAAAATCGAA-----GCATACGA	768
Qy	sGlnGlyIleTrpLysGluAlaMetSerAlaAsp-----GlnAlaThrArgIl	304
Qy	:	
Db	AGACCGGATAATCTTGAATAATACATAGCGGACATCACAGGACTCGAAAACCTCCGGGT	804
Qy	eLysTyrAlaSerLysTyr---AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMe	320
Qy	:	
Db	AAAAACAAAGCCCTCTATCTCGGGCTCACCGAAGAAAAACACGAAGACCGGGCGGTG	864
Qy	tAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArgAlaPh	339
Qy	:	
Db	GACAAAGCGCCTTTTCGACACCTTTGTGGAACCTCGGTGTTTACACACGCTTCATATCTC	924
Qy	eAlaAspTrp-----IleArgLysAsnGlyLysSerAlaValTyr-----	359
Qy	:	
Db	CTTTGATTCTACACAGGCGACGAACTCGATAAAGATT--CGAAGTAGTCTACAACCTGG	984
Qy	-----GlyAspValLeuSerSerLeuGlyLysAl	373
Qy	GTACGATCCTTACCTGTTCTGTTCCGGAGGCGACATCTCAACCGATCCCAAAACC	1043
Qy	aTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSergLuthrLeuPheGl	382
Qy	:	
Db	CACACACGAGAATCAGAGAAGTCAAAGAAATGGTCAAGCCCTTCACAAAACGGTATAG	1103
Qy	y-----GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAs	402
Qy	:	
Db	GTGTGATTATCGACATGGTGTTCCTCCCTCACACTACGGTATAGCGAACTCTCGCGTTCG	1163
Qy	nProAspAlaHisAlaGlyIleLeuLysSerLeuAspLys-----	419
Qy	:	
Db	ATCAGACGGTCCCGTACTACTACTTCTACAGATCGACAGACAGGTCCTATTTCGAACGAA	1223
Qy	-----TyrLysAspTyrLeuProSerLeuAs	434
Qy	:	
Db	CGGGATGGTAACTGTCATCGCAAGCAAGACCCATGATGAGAAAATTCATGTCGATA	1283
Qy	pArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleProAla-AspLysL	442
Qy	:	
Db	CCGTCACTACTGGGTAA-----GGAGTATCATATAGACGGATTTCAGGTTCGATCAGA	1343
Qy	eUpProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyr-	462
Qy	:	
Db	GGTCTCTATCGCAAAAAAGACAATGCTCGPAAGTCGAAAGAGCTC	1397
Qy	-----AlaA	482

[illegible]

```
Db 1442 TTCATAAATCGATCAACTATCATCTCTACGGCGAACCGTGGGGTGATGGGAGCAC 1501
Oy 483 spPheValpheaSpLysSer-----ValValProTyrSerAspLysp 497
Db 1502 CGATCAGGTTTGGAAAGACCGATGTCGCCGGGCACACAGCTGGCAGCTTTCAACGATGAGT 1561
Oy 497 heHisAlaMetLeuLys-----s 503
Db 1562 TCAGAGACCAATAAGGGGTTCCGGTTTCAACCCGAGCGTCAAGGGATTCTGTCATGGGAG 1621
Oy 503 erMetAspLysGlu-----L 508
Db 1622 GATACGGAAAGCAAAACCAAGATCAAAAGGGGTGTTGTTGGAAGCATATAACTACGACGGAA 1681
Oy 508 ysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer----- 522
Db 1682 AACTCATCAAAAGTTCCCTCCCTTGATCCAGAGAAACTATTAACCTACCGACGGTGTACAG 1741
Oy 523 -----LysSerValIleAlaAla-----AlaArgAlaI 532
Db 1742 ACAACACACACTGTGGGACAGAACTACCTTGGCGCCAAAGCTGATAAGAAAAGGAAT 1801
Oy 532 leGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheA 552
Db 1802 GCACCGAAGAGAACTGAAAACGCCGCAAACTGGCTGGTCCGATACTTCTCACTTCTC 1861
Oy 552 laGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMeta 572
Db 1862 AAGGTGTTCTTCTCCACGGAGGCGAGACTCTCGCAGGACGACGAATTTCAACGACA 1921
Oy 572 rgMetSerTyrGly-----SerIleLysGlyTyrGluProGlnAspGlyAlaTrp- 588
Db 1922 AC---TCCTACAACGCCCTATCTCGATAAAGCGTTCGATTACGAAAGAACTTCAGT 1978
Oy 589 -----TyrAsnTyrHisThrGlyLysGlyVal-----LeuGluLysGlnA 603
Db 1979 TCATAGACGTGTTCAATTACCAC-----AAGGGTCTCATAAACTCAAGAAAGAAC 2029
Oy 603 spProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysA 623
Db 2030 ACCCT-----GCTTTACGGCTGAAAA 2050
Oy 623 snTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn----- 640
Db 2051 AC-----GCTGAAGAGATCAAAAACACCTCGGAATTTCTCCCGGGGGGAGAA 2098
Oy 641 -----AsnAspIleThrGlyLysAsnSerGlySerPro---ValP 653
Db 2099 GAATAGTTGCGTTTCATGCTTAAAGACACACCGAGGTGGTGTATCCCTGGAAAGACATCGTGG 2158
Oy 653 heAspLysAsnGlyArgLeuIleGlyLeuAlaPhe-----AspGlyAsnTrp---- 668
Db 2159 TGATTTACAAATGGAAACTTAGAGAGACAACTAACTCCAGAGAAAGAAATGGAAATG 2218
Oy 669 -----GluAlaMetSerGlyAspIleG 676
Db 2219 TGGTTGTGAACAGCCAGAAAGCCGGAACAGAGTATAGATAAAACCGTCCGAGGAACATAG 2278
Oy 676 luPheGluPro 679
Db 2279 AACTCGATCCG 2289
```

## RESULT 12

```
US-09-134-001C-195
; Sequence 195, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
```

```
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 195
; LENGTH: 2667
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-195

Alignment Scores:
Pred. No.: 0.059 Length: 2667
Score: 110.00 Matches: 112
Percent Similarity: 34.72% Conservative: 89
Best Local Similarity: 19.34% Mismatches: 178
Query Match: 2.96% Indels: 201
DB: 4 Gaps: 28

US-10-008-355-2 (1-712) x US-09-134-001C-195 (1-2667)
Oy 183 AsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPheAla 202
Db 100 AGTAATGGTTACTTTAAACCATCTCAAGATAAATCTAAAGAGGCATATACAAATTGTAATT 159
Oy 203 ProProSerValGlyLysPheGlyAspThrAspAsnTrpMetTrpProArgHis 222
Db 160 CCGCCC-----CCTAATGTA 174
Oy 223 ThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyr 242
Db 175 ACAGGTAATTTACATTTTAGTCATGTCATGG----- 204
Oy 243 SerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyr 262
Db 205 -----GATACACTTTTACAAGATATAAATACTAGAATGAAAAGAAATCGAAGGATAC 255
Oy 263 LysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThr 282
Db 256 GAT-----ACTTTATATCTGCTGGTATG----- 279
Oy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302
Db 280 -----GATCATGCTGGTATAGCAACTCAAGCAAAAGGTTGAAGCAAAACTT 324
Oy 303 IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyr 322
Db 325 AATGAACAAAGGTATATCTAGACACGACTTTAGGAAGAGAA----- 363
Oy 323 AlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsn----- 335
Db 364 -----AAATTTTACAGCAAGCATGGGATTGGAAAGAGAGATGCAACATTTATTAGA 417
Oy 336 -----SerIleGlyMetAsnArgGlyLeuAlaArg-----LeuAsp 347
Db 418 CAACAATGGGCTAAACTTGGCTTAGGTTAGATTATAGTAGAAGACGGCTTTACGCTTAGAT 477
Oy 348 VallleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLys----- 365
Db 478 ---GATGGTTTAAAGTAAAGCTGTAAAGAAAGTATTTGTTGATTTATATAATAAGGTATT 534
Oy 366 -----AsnGlyLysSerAlaValTyrGlyAsp-----ValLeuSerSer 378
Db 535 ATTTATCGTGTGAAGAGAAATTTAACTGGGACCCCTATAGCTAGAACAGCTTTATCAGAT 594
Oy 379 LeuGluLysAlaTyrLys-----GluGlyAlaLysAlaAsnArgGluMetThrTyr--- 395
Db 595 ATTGAAGTCATTTCATGAAGATGTCACAGGTGCTTTTATCATCTTTAAATATATCCTTATGTT 654
Oy 396 -----LeuSerGluThrLeuPheGlyGlyThr 404
Db 655 GATGGGAATGGATATATAGAAATAGCTACTACACCTCTCTGAGACTATGTTAGGTGACACT 714
```

```
QY 405 GluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla 424
Db 715 -----GCTATTGTTGTAATCCAAAT----- 735
QY 425 GlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLys 444
Db 736 -----GAGCAAGAGTAAAGATGATGTTGTTAA----- 765
QY 445 ValLeuProAlaMetLeuAspIleValArgArgIlePro-----AlaAspLysLeu 462
Db 766 -----ACAGTTATATTACCTATTGTTAGGAGAGAGTACCTATTGCTGATGATAC 819
QY 463 ProAspIle----- 465
Db 820 GTTGATATTGAATTTGTTTCAGGAGCAATGAAAGTTACTCCGCCACATGATCCTAATGAC 879
QY 466 -----PheLysAsnValIle-----AspLysLysPheLys 475
Db 880 TTTGAAATTTGGTCAAGACATCAATTTAGAAATATTATAGTCATGGATGATGATGTTAAAG 939
QY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db 940 ATGAACGATAAA-----GCTGAT 957
QY 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIle---Glu 514
Db 958 AAATACAAAGGGATGGATAGATTGTTGTTAGAAATCAGCTTGTAAAGACTTAAAGAA 1017
QY 515 LysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGluAla 534
Db 1018 CAAGATCTTGTATAAATTTGAAGAATACATACATTCAGTAGGACATTCCTGAACGTTCT 1077
QY 535 AspAlaMetAlaAsnAlaIle----- 541
Db 1078 GGTCCANTAGTAGACCTTATTTGCTCTACACAATGGTTGTAAATAATGAACCTTTGGCA 1137
QY 542 -----AlaIleGlu-----LysGlyLysArgLeuPhePheAlaGlyLeuArg 555
Db 1138 CAAAGAGCTTTGGATATCAAAATACAAAGATAGAATTTGACTTTTCCAGGAGTAGATT 1197
QY 556 GluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMet-ArgMetSerTyr 575
Db 1198 GAAATACATTCATCCTGGATGAGAAATTTAGAGATTGGAGGATTCGCGTCAACTA 1257
QY 575 rGlySerIleLysGlyTyrGluProGlnAspGlyValAlaTyrAsnTyrHisThrGlu 595
Db 1258 TGGTGGGGCCCAAAATACCT-----GCTTGGTAT---CATAAAGATACAGG 1301
QY 595 YLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln---GluAs 614
Db 1302 TGAAGTCTTTGTTGGTGAAGAAGCACCTGAGACATCGAATTTGGATACAAAGATGAGGA 1361
QY 614 nIleLeuAspLeuPheArgThrLys-AsnTyrGlyArgTyrAlaGlu---AsnGlyGlnL 633
Db 1362 TGTATTAGATACATGGTTCTCAAGTCAGCTTTGCCGTTTTCACACATTAGATGGCCTGA 1421
QY 633 euHisIleAlaPheLeuSerAsnAsnAspIle-----T 644
Db 1422 TACAAATGCTGATGATTTTAAACGCTTATATCCGACAAATGCTATTACTGCTTATGA 1481
QY 644 hrGlyGlyAsnSerGlySerProVal---PheAspLysAsnGlyArgLeuIle 660
Db 1482 TATCATTTTCTCTCGGGTAGCCCGTATGATTTTCCAGGATTTAGAAATTTACTG 1534

RESULT 13
US-08-446-137B-1
; Sequence 1, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowen, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
```

APPLICANT: Laurence, Oliver S.  
APPLICANT: Dugleby, Clive J.  
TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED  
TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, version #1.30  
CURRENT APPLICATION DATA: US/08/446,137B  
APPLICATION NUMBER: US/08/446,137B  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: McWaters, David D.  
REGISTRATION NUMBER: 33,963  
REFERENCE/DOCKET NUMBER: 100084.406  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3279 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Peptostreptococcus asaccharolyticus  
STRAIN: 1018  
FEATURE:  
NAME/KEY: exon  
LOCATION: 103..3186  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /codon\_start= 280  
OTHER INFORMATION: /product= "mature protein L"  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /number= 1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 103..3186  
OTHER INFORMATION: /codon\_start= 103  
OTHER INFORMATION: /product= "immature protein L"  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 280..3183  
OTHER INFORMATION: /codon\_start= 280  
OTHER INFORMATION: /product= "mature protein L"  
FEATURE:  
NAME/KEY: misc\_signal  
LOCATION: 208..279  
US-08-446-137B-1

Alignment Scores:  
Pred. No.: 0.0812 Length: 3279  
Score: 110.00 Matches: 118  
Percent Similarity: 31.59% Conservative: 63  
Best Local Similarity: 20.59% Mismatches: 206  
Query Match: 2.96% Indels: 186  
DB: 4 Gaps: 29

US-10-008-355-2 (1-712) x US-08-446-137B-1 (1-3279)



Qy 29 TrpLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThr 48  
|||||  
Db 1861 TGGTTATTAAAGATGCTAAAGAGAACCAATCAAGAATTAAGAAGACGAGCAATCACT 1920  
Qy 49 LeuProLeuAspSerLeuTyrSerPheAspLysProSerIleAlaAsnAlaValValIle 68  
|||||  
Db 1921 -----TCTGATTATTATC-----TTACGCTTAATCAATAAGCAAAAACAGTT----- 1962  
Qy 69 PheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88  
|||||  
Db 1963 -----GAAGCGGTGAAGCATTTAAAGAACGAAATCTTA-----AAAGCA 2001  
Qy 89 HisCysGlyTyrGlyAlaIleGlnSerGlnThrValAspHisAspTyrLeuArgAsp 108  
|||||  
Db 2002 CACGCTGA-----GAAGAACACACAGAA-----TTAAAGAT 2034  
Qy 109 GlyPheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSerValLysTyr 128  
|||||  
Db 2035 GGATATGCAACATATGAAGAAGACGACGACGCTAAAGAGCTTTGAAAAATGATGAT 2094  
Qy 129 LeuArgLysIleValLysValThrAspLysValGluGlyGln-----LeuLys 144  
|||||  
Db 2095 GTTAAACACGATACGAATAATAGTTCAAGGTGCAGACGGAAGATCTACTATGTATTAAAG 2154  
Qy 145 ---GlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln----- 161  
|||||  
Db 2155 ATTGAAGTTGCAGAGAGAGAACCGAGTGAAGACATCTCCAGAGTTCAAGAGGTTAC 2214  
Qy 162 -----GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCys 175  
|||||  
Db 2215 GCAACTTACGAAGACGACGACGACGCTAAAGAGCATTTAAAGAGATAAAGTT--- 2271  
Qy 176 IleValGluProPheTyrSerAsnAsnGluTyrPheLeuIle----- 189  
|||||  
Db 2272 -----AACAATGCATACGAAGTAGTTCAAGGTGCAGACGGAAGA 2310  
Qy 190 ValTyrAspValPheLys---AspValArgMetValPheAlaProProSerSerValGly 208  
|||||  
Db 2311 TACTACTATGATTAAANAATCGAAGATAAAGAGATGAACACCGAGGTGAAGAACCGGC 2370  
Qy 209 LysPheGlyGlyAspThr---AspAsnTrpMetTrpProArgHisThrGlyAspPheSer 227  
|||||  
Db 2371 GAAACCCAGGAATCACAAATCATGATGATGATTTAAAGATGCTAAAGAGAC----- 2424  
Qy 228 ValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLys 247  
|||||  
Db 2425 -----GCAATCAAGAAATTAAGAGAACGAGCAAGCAATCAGTTCTGAC----- 2463  
Qy 248 ProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspTyr 267  
|||||  
Db 2464 -----ATATACTTTGATGCAATC----- 2481  
Qy 268 AlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGlu 287  
|||||  
Db 2482 -----AACAAAGCAAAACACAGTAGTAGAGCGGTAGAA 2511  
Qy 288 AspArgIleGluAsnGlu-----AsnAsnProArg 297  
|||||  
Db 2512 ---CGGTTAAAGAGCAAAATCTTTAAAGCACACGCTGAAAAACCGCGCAAAACCCGGA 2568  
Qy 298 IleGluVal-----ArgGlyIleLysGlnGlyIleTrpLysGlu----- 310  
|||||  
Db 2569 ATCACAATTGATGATGTTATTAAAGAATGCTAAAGAAGCTGCAATCAAGAATTAATAA 2628  
Qy 311 -----AlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGln 328  
|||||  
Db 2629 GAACGAGGAATCACTGCTGAATATCTATTCAACTTAATCAACAAAGCAAAACACAGTAGAA 2688  
Qy 329 SerAlaAsnTyrTrpLysAsnSerIle-----GlyMetAsn 340  
|||||  
Db 2689 GCGGTAGAAATCAATTAAGAGCAAAATCTTTAAAGCACACGCTGAAAAACCGCGCAAAAC 2748

Qy 341 ArgGlyLeuAlaAlaArgLeuAspValIleGlyArgLysArgAlaGluArgAlaPheAla 360  
|||||  
Db 2749 CCAGGAATC-----ACAATTGAT 2766  
Qy 361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380  
|||||  
Db 2767 GAATGGTTATTAAAGACGCTAAA-----GAA 2793  
Qy 381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400  
|||||  
Db 2794 GATCAATTAAGAA-----TTAAAGAGCAGCAAGTAATTACTTCGACATATAC 2841  
Qy 401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeu----- 416  
|||||  
Db 2842 TTTGATGCTATCAACAAAGCAAAACTATTGAAGCGGTAGAAGCATTTAAAGATCAAAATC 2901  
Qy 417 -----AlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434  
|||||  
Db 2902 TTAAGGCTCATATAAAGATGAAGACCAAGTAAAAAACCCAGGTGAAGACAAAAACCA 2961  
Qy 435 LysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArg 454  
|||||  
Db 2962 GAAGATAAAACCAAGGTGAAGATAAAAA----- 2991  
Qy 455 ArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPhe 474  
|||||  
Db 2992 -----CCAGAAGACAAAAACCTGCTGAAGATAAAAAAACCCAGACAAAAA--- 3039  
Qy 475 LysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSer 494  
|||||  
Db 3040 CCAGTAAACAGATAA----- 3057  
Qy 495 AspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGlu 514  
|||||  
Db 3058 -----GATTCTCCAATTAAGAAAAAAGCTTAATA----- 3090  
Qy 515 LysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAla 534  
|||||  
Db 3091 -----CCAAAGCTGGTAGCGAAGCTGAATCTTTAACATTAGCAGCAGCAGCTTTATCA 3144  
Qy 535 AspalameAlaAsnAlaTyrAlaIleGluLysGlyLys 547  
|||||  
Db 3145 ACAGCAGCAGGTGCTTACGTTTTCACCTTAAAAAACCTAAA 3183  
RESULT 14  
US-08-961-527-139/c  
; Sequence 139, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders



; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 139:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 15363 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-961-527-139

Alignment Scores:  
 Pred. No.: 1.14 Length: 15363  
 Score: 109.00 Matches: 124  
 Percent Similarity: 29.03% Conservative: 83  
 Best Local Similarity: 17.39% Mismatches: 222  
 Query Match: 2.93% Indels: 284  
 DB: 4 Gaps: 30

US-10-008-355-2 (1-712) x US-08-961-527-139 (1-15363)

Qy 210 PheGlyGlyAsp-----ThraspAsnTrpMetTrpProArg 221  
 Db 10583 TATGGAAATGATGCTGGGTGGGGGAAACACACACCGCTTCATGGACTTTTGACCGT 10524  
 Qy 222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAsp-----AsnArgPro 239  
 Db 10523 GACAAACGCTGCTATGCTGGACAGTTATCTCGGACAGTACGGACTATATTGGTGAACCT 10464  
 Qy 240 AlaGluTyrSerLysAspAsnLys---ProTyrLysProValTyrPheAlaAlaValSer 258  
 Db 10463 ACACCATGGCACAACCAACCAACCTCTGTTAAGAGCTCTTACTTTGGTATCGTAGAT 10404  
 Qy 259 MetGlnGlyTyrLysAlaAspTyrAlaMetThrIleGlyPheProGlySerThrAsp 278  
 Db 10403 ACAGCCGGCATTCACAAACATGACTTCTATCTACCAAAACCAATGGGTTCTGTAAAG 10344  
 Qy 279 Arg-----TyrLeuThrSerTrpGlyValGluAspArg----- 289  
 Db 10343 AAGAAACCGATGGTACACCTTCTCCCTCAGTGGAAACCAAGAAGATTAGCATCC 10284  
 Qy 290 -----IleGluAsnGluAsnAsnPro 296  
 Db 10283 AAAGTAGCTGACTCAGAAGSTAAGATTCCAGTTCGTGCTTATTCGAATGCTTCTAGTGA 10224  
 Qy 297 ArgIleGluValArgGlyIleLysGlnGlyIle----- 307  
 Db 10223 GAATGTCTTGAATGAAATCTCTGGTCTTAAGACTTTCATATAAAACAAACACAGC 10164  
 Qy 308 -----TrpLys----- 309  
 Db 10163 GATGGCGGACTTACCAAGAAGTGCRAATGCTAATGAACCTTATCTTGATGGAAGTT 10104  
 Qy 310 -----GluAlaMetSerAlaAspGlnAla----- 317  
 Db 10103 GCCATCAACAGGTACCTTGGAAAGCAATTCGTGATGAATCTGCGCAAGAAATGCT 10044  
 Qy 318 ---ThrArgIleLysTyrAlaSerLys----- 325  
 Db 10043 CGATATAAGATTACGACTGCTGGTAAGCCAGCGGAGTTCGTCTTATTAAGAAGACCAT 9984  
 Qy 326 -----TyrAlaGlnSerAlaAsnTyrTrp 333  
 Db 9983 CGGATTGCAGCAGATGGAAAGACTTGACTTACATCTACTATGAATGTGTCACAGCCAG 9924  
 Qy 334 LysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGly----- 350  
 Db 9923 GGGAAATGGTTCCAACTGCTTAATAATCTGTTCCCAATTCATGGTCGCAAGGTCAA 9864  
 Qy 351 -----ArgLysArgAlaGluGluArgAlaPhe 359  
 Db 9864 GGGAAATGGTTCCAACTGCTTAATAATCTGTTCCCAATTCATGGTCGCAAGGTCAA 9864

Db 9863 CTGTCGGTGTAGATAACGGAGAACACCCAGCCGCTGAACGCTATATAGCCGCAACAGCAT 9804  
 Qy 360 AlaAspTrpIleArgLys-----AsnGlyLysSer----- 369  
 Db 9803 GGTTCCTGGATTCGTAAGCATTTAATGTAAAGTGTGGCATTTGTCAAATCAACTGAA 9744  
 Qy 370 -----AlaValTyrGlyAspValLeuSerSerLeuGluLysAla 382  
 Db 9743 CAAGCAGGGAAATTCACCCCTGACTGCCACCTCTGATCTCTTGAATCGAACCAAGTCACT 9684  
 Qy 383 TyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGly 402  
 Db 9683 GTCTTTTACTGTAAAGAGAGGACAGAGAACTGTTTGT----- 9642  
 Qy 403 GlyThrGluVal----- 406  
 Db 9641 GGCACAGAAAGTGCACCAAGTACAGACCATTTATTGGAGAGGCACCTGAAATGCCTACCAC 9582  
 Qy 407 ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIle 426  
 Db 9581 GTTCCGTTTGTATACAGTGTAGCGGTGACAGAACCTCTGTAACTGGTCTTTCAGTA 9522  
 Qy 427 -----LeuLysSerLeuAspAspLysTyrLysAspTyr 437  
 Db 9521 CATGTGACCAAGCCCTGGTATTGTAAAGGTGAAGATGTGGTGACGGA---CGAAGATA 9465  
 Qy 438 LeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIle 457  
 Db 9464 GAAGCTCGTGTAGAAGTATTGCTCTTAATACAGAGCTACCAAGTTGTGAACGTTATGCT 9405  
 Qy 458 ProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAsp 477  
 Db 9404 CCAATACTGACTTGAATTCCTGTAGACAAATCTGTTTCTCTATGTTTGTGATGCGAAT 9345  
 Qy 478 ThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLys--- 496  
 Db 9344 GTTGAAGATATGAA-----GTGGACAAGTGGGAGATTGCCGAGAAGATAAAGCT 9294  
 Qy 497 -----Phe 497  
 Db 9293 AAGTTAGCAATTCACAGGTTCTCGTATTCAAGCGCCGGTATTATTAGAGGTCAACCAATT 9234  
 Qy 498 HisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspPro 517  
 Db 9233 CATGCAACCTTGTGGTGAAGAAGCAATCTCGCGCACCTGCAGTA-----CCA 9183  
 Qy 518 AlaValGluLeuSerLysSerValIle----- 526  
 Db 9182 ACTGTAAACGGTGTGGTGGTGGCAGTAAACAGTCTTACTAGTCAAAAACCAATGCAATAC 9123  
 Qy 527 -----AlaAla 528  
 Db 9122 CGCACTCTTGTGAGCTAAGTTGCCAGAAGTCACAGCAAGTGTATAAATGCACT 9063  
 Qy 529 AlaArgAlaIleGlnAlaAspAlaMetAlaAsn----- 539  
 Db 9062 GTTACAGTTCTTCAAGCAAGCGCA---GCAACGCGCATCGCTGCGAGCATCTTTATTTCAG 9006  
 Qy 540 -----AlaTyrAlaIle-----GluLysGlyLysArg 548  
 Db 9005 CCTAAAGATGGTGGCCCTCTTCAAAACCTTATGCAATTCCTTGAAGACGCGCAAAA 8946  
 Qy 549 LeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAsp----- 566  
 Db 8945 ATTGCTCATTGAGCTTG---CAAGTGGAAAAGCTGACAGTCTCAAGAGAACCAACT 8889  
 Qy 567 AlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGly 586  
 Db 8888 GTCAAATGTGGTTCGAGTCACTAT-----CAAGATGGA 8853  
 Qy 587 AlaTrp-----TyrAsnTyrHisThrThrGlyLysGlyValLeu 599  
 Db 8852 ACGCAAGCTGTATTACCAGCTGATAAAGTAACTTCTCTACAAGTGGTGAAGG----- 8799

Qy 155 LysAlaGlnGluValCysGlnLeuLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeu 174  
 Db 4015644 -----CTGCCAAGCCGCCGCCGGCG----- 4015621  
 Qy 175 CysIleValGluProPheTyr-----SerAsnAsnGluTyrPheLeuIleVal 190  
 Db 4015620 -----CGGCCGTCGTGGGAGTGGCGCGCTACCCACAGAGCGTGCCTTCACCGCGGCC 4015570  
 Qy 191 TyrAspValPheLysAspValArgMetValPheAlaProProSerSerValCylLysPhe 210  
 Db 4015569 TACGACCTCTG-----GACGTC-----TTGCCCGCGTCCGCCGGAG----- 4015534  
 Qy 211 GlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe--- 229  
 Db 4015533 -----CCGCGCGCCACCGGCATATCACCCAGATGATC----- 4015501  
 Qy 230 -----ArgValTyrAlaGlyAlaAspAsnArgPro 239  
 Db 4015500 GAGATGATCAGCGCCCTGATCCAGCGCGCCACGCGTATACCGGTGGCGGTGAC----- 4015447  
 Qy 240 AlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSer--- 258  
 Db 4015446 -----GTTTACTTCGACGTTCTGAGGTAC 4015423  
 Qy 259 -----MetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhe 273  
 Db 4015422 CCGGAGTACGCCAGTTATCCGGTFCACAAGATCGATCCATCAGGCGGAAGGTGTG 4015363  
 Qy 274 --ProGlySerThrAsp---ArgTyrLeuThrSerTrpGlyValGlu-AspArgIleG1 291  
 Db 4015362 GCTCGCGCGAAGCGGACCGACCGACTTCACCTTTGTGGAAGGCGCAAAACCGCGGTG--- 4015305  
 Qy 291 uAsnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAl 311  
 Db 4015304 -ACCGTCGTGGCGGACCGCGTGGGCGCGCGGCTCGGCATCTCG----- 4015254  
 Qy 311 aMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAs 331  
 Db 4015253 -----AATGCTCG 4015246  
 Qy 331 nTyrTrp-----LysAsnSerIleGlyMetAsnArgGlyLeuAlaAr 345  
 Db 4015245 GCAATGGCTCGCAGCTATCTCGGCGCGGAATTCGATATCCATTGC--GGTGAATGGATT 4015187  
 Qy 345 gLeu-----AspValIleGlyArgLysArgAlaGluAlaPheAl 360  
 Db 4015186 AGTCTTCCCGCATCATGAGAACGAATATGCCAGCGCGCGCGCGGTGATGGGTTCGC 4015127  
 Qy 360 a-----AspTrpIleArgLysAsnGlyLys-----SerAlaValTy 372  
 Db 4015126 CCGCTACTGGCTGCACATGGCTGGGTGACGATGGCGGGGAGAGATGAGCAAGTCGCT 4015067  
 Qy 372 rGlyAspValLeuSer-----SerLeuGluLysAlaTy 383  
 Db 4015066 GGGCAACGCTTCTCCATGTCGGCGGATGTTGCAGCGGGTGGCGCGGTGATGCTGCTTA 4015007  
 Qy 383 rLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGlnThrLeuPheGlyG1 403  
 Db 4015006 TTACTTGGGAGCGCGGCACACTACCGCGGTGATGCTCGAGTCTCCGAGACC----- 4014958  
 Qy 403 yThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHi 423  
 Db 4014958 ----- 4014958  
 Qy 423 sAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspAr 443  
 Db 4014957 -----GCTATGAGGATCGCGTGAAGACCTATGTCGG----- 4014925  
 Qy 443 gLysValLeuProAlaMetLeuAspIleValArgArgIle-----ProAl 459  
 Db 4014924 -----CTAGAGGACTTCTGCACCGCGTTCGACCCCGGTGGCGCGCTCTGCCCGCG 4014872  
 Qy 459 aAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGluValAspThrLy 479

```
Db 4014871 CGATCCGACCCCA----- 4014859
Qy 479 sLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAl 499
Db 4014859 ----- 4014859
Qy 499 aMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaVa 519
Db 4014858 ----- 4014824
Qy 519 lGluLeuSerLysSerValIleAlaAlaAlaAraGAlaIleGlnAlaAspAlaMetAlaAs 539
Db 4014823 TCCGATCCGGCTCGCGAGATTCCACCAGTCCGGGG----- 4014787
Qy 539 nAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrPr 559
Db 4014786 -----GAAGGCAACCGGCACCTCGATCCGGCGACACGAC----- 4014751
Qy 559 oGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLy 579
Db 4014750 -----GGGGGCTTCGGAAGCGCTAGCGCGATTCCGGCGCATGATG-----GGCATCCT 4014704
Qy 579 sGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHisThrThrGlyLysGlyValLe 599
Db 4014703 GGCTGTGACCCCTCGACGAGCGCTGG----- 4014676
Qy 599 uGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu-- 618
Db 4014675 -GAATCCCGAGACGAAACCTCGCGACGCTGGCGCGCTCGATGCTGCTCCAGGCTGA 4014617
Qy 619 -PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLe 638
Db 4014616 ACTACAGAATCGGGAAGCGCGAGCAGCGCAACTGGCGCTC----- 4014571
Qy 638 uSerAsnAspIleThrGly-----GlyAsnSerGlySerProValPheAspLysAs 656
Db 4014570 -----GCCGAGGATCCGGGTCCGGTGAAGAGAGCGCGCATCGAGGTACCCGAC----- 4014520
Qy 656 nGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGl 676
Db 4014519 -----ACCGCCGAGGGCCACAGTGGTCTGCTGGTGGTGACACCAA 4014476
Qy 676 uPheGluProAspLeuGlnArg 683
Db 4014475 GTAGATGCCCGGTAACCTCGG 4014454
```

Search completed: May 23, 2003, 13:33:17  
Job time : 4891.71 secs



GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 10:12:58 ; Search time 186.87 Seconds  
(without alignments)  
5031.128 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MQMKLSILLGALLGASG.....LFMDKWGCPRLIQELKLI 712

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame_plus_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10008355/runat_16052003_110406_2649/app_query.fasta_1.1230
-DB=PublishedApplications_NA -OFFT=fastap -SURFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10008355 -CGN_1_1_132 -runat_16052003_110406_2649
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database :

```
PublishedApplications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3719	100.0	2139	9 US-10-008-355-1	Sequence 1, Appli
2	133	3.6	2652	10 US-09-815-242-9284	Sequence 9284, Ap
3	127	3.4	2652	10 US-09-815-242-9586	Sequence 9586, Ap
4	116.5	3.1	15424	9 US-10-143-186-1	Sequence 1, Appli

5	116.5	3.1	15424	9	US-10-143-186-2	Sequence 2, Appli
6	116	3.1	3598	10	US-09-925-301-170	Sequence 170, App
7	113.5	3.1	2856	10	US-09-815-242-6368	Sequence 6368, Ap
8	112.5	3.0	10758	7	US-08-781-986A-221	Sequence 221, App
9	110.5	3.0	2082	9	US-09-769-787-197	Sequence 197, App
10	110.3	3.0	15450	9	US-10-127-391-1	Sequence 1, Appli
11	110	3.0	2319	9	US-10-121-032-14	Sequence 14, Appl
12	110	3.0	2319	9	US-10-093-037-14	Sequence 14, Appl
13	110	3.0	2538	9	US-10-027-806-61	Sequence 61, Appl
14	110	3.0	2538	9	US-10-034-623-61	Sequence 61, Appl
15	110	3.0	2538	9	US-10-027-801-61	Sequence 61, Appl
16	110	3.0	8930	9	US-10-192-584-1	Sequence 1, Appli
17	110	3.0	32998	9	US-10-027-806-1	Sequence 1, Appli
18	110	3.0	32998	9	US-10-034-623-1	Sequence 1, Appli
19	110	3.0	32998	9	US-10-027-801-1	Sequence 1, Appli
20	109	2.9	6702	9	US-09-769-787-209	Sequence 209, App
21	108	2.9	3825	10	US-09-974-300-1925	Sequence 1925, Ap
22	108	2.9	11871	10	US-09-070-927A-171	Sequence 171, App
23	107	2.9	38155	9	US-10-114-170-79	Sequence 79, Appl
24	106.5	2.9	1572	10	US-09-815-242-7820	Sequence 7820, Ap
25	106.5	2.9	5761	9	US-09-928-847B-26	Sequence 26, Appl
26	106	2.9	2472	9	US-10-100-049-7	Sequence 7, Appli
27	105	2.9	3358	9	US-10-100-049-20	Sequence 20, Appl
28	104.5	2.8	2610	10	US-09-815-242-6526	Sequence 6526, Ap
29	104.5	2.8	3513	9	US-09-938-842A-862	Sequence 862, App
30	104.5	2.8	5308	10	US-09-070-927A-444	Sequence 444, App
31	104.5	2.8	9821	7	US-08-781-986A-470	Sequence 470, App
32	104	2.8	3731	10	US-09-782-874-1	Sequence 1, Appli
33	104	2.8	7639	9	US-09-971-536-28	Sequence 28, Appl
34	104	2.8	14121	9	US-09-802-840-31	Sequence 31, Appl
35	104	2.8	14121	9	US-09-920-033-3	Sequence 3, Appli
36	103.5	2.8	1582	9	US-09-470-526-1	Sequence 1, Appli
37	103.5	2.8	1959	9	US-09-470-526-8	Sequence 8, Appli
38	103.5	2.8	3640	9	US-09-292-758-144	Sequence 144, App
39	103.5	2.8	4392	9	US-10-127-391-3	Sequence 3, Appli
40	103	2.8	1191	10	US-09-801-368-137	Sequence 137, App
41	103	2.8	2387	10	US-09-974-300-643	Sequence 643, App
42	102.5	2.8	4142	9	US-09-925-299-138	Sequence 138, App
43	102.5	2.8	4142	10	US-09-925-299-138	Sequence 138, App
44	102.5	2.8	6142	9	US-10-014-436-1	Sequence 1, Appli
45	102.5	2.8	7053	10	US-09-070-927A-78	Sequence 78, Appl

#### ALIGNMENTS

```
RESULT 1
US-10-008-355-1
; Sequence 1, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Barbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-1
```

Alignment Scores:  
Pred. No.: 0  
Score: 3719.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
Length: 2139  
Matches: 712  
Conservative: 0  
Mismatches: 0  
Indels: 0

DB: 9 Gaps: 0  
US-10-008-355-2 (1-712) x US-10-008-355-1 (1-2139)  
QY 1 MetGlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGly 20  
DB 1 ATGCAAAATGAAATTAATAAGATATTCTTCGAGACAGCCCTGCTGTGGGTGCTTCAGGG 60  
QY 21 ValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlnLeuAsnGlnGlnAsnLeuAsp 40  
DB 61 GTAGCCAAAGCCGACAAAGGATGTGGCTCTCAACGACTCAATCAGGAGAATCTGGAT 120  
QY 41 ArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyrSerPheAspLysPro 60  
DB 121 CGAATGCGTGAGCTCGGCTTTACGCTCCCGTTGGATTGCGTCTACAGTTTCCACAAAGCCG 180  
QY 61 SerIleAlaAsnAlaValIlePheGlyGlyGlyCysThrGlyIleThrValSerAsp 80  
DB 181 TCCATTTGCCAATGCGGTGTTATCTTCGGTGGCGGATGTACCGGTATCACAGTCTCCGAT 240  
QY 81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100  
DB 241 CAGGCGCTGATCTTTACCAACCACTCGCGATACGGTGTATCCAGAGCCAAAGCACG 300  
QY 101 ValAspHisAspTyrLeuArgAspGlyPheValSerArgThrMetGlyGluLeuPro 120  
DB 301 GTGATACAGACTATCTCGCGCATGGTTCGTTCTCGCAGATGGGTGAGCAGCTTCG 360  
QY 121 IleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLysValGlu 140  
DB 361 ATTCGGGTCTTTCCGTAAGATATCTGCGCAAGATCTGGAAGGTAAACGACAGGTAGAA 420  
QY 141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnIleValCys 160  
DB 421 GGACAGCTCAAGGTATCAGTCAGAGATGAGCGCTCTGCCAAGCTCAGCAGGTATGC 480  
QY 161 GlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180  
DB 481 CAAGAACTGGCCAAAAGAAATGCAAGACGAGACGAGAACCAACTCTGCATCGTAGAGCTTTC 540  
QY 181 TyrSerAsnGlnTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200  
DB 541 TATTCCAAACAAGAAATACTCTCATCGTACCATGATTTCAAGGACGTTTCGTATGGTA 600  
QY 201 PheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpPro 220  
DB 601 TTTGCTCTCCACGCTGTAGTAGTAAGTTCGAGCGCATACGGACAACTGGATGTGCGCG 660  
QY 221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAla 240  
DB 661 CGTCACAGGGCGACTTCAGCGTATTCGCGTGTATGCGGTGTCGCGACAAACCGCGCGCC 720  
QY 241 GluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetCln 260  
DB 721 GAATACAGCAAGGACAATAAACCTTAAGCCCGTTTACTTCGCTGCGGTATCCATGCAA 780  
QY 261 GlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyr 280  
DB 781 GGCTACAAGGCTGACGACTATGCCATGACCATCGGTTTCGCGGCAGTACGGATCGCTAC 840  
QY 281 LeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluVal 300  
DB 841 CTCACTCTTTGGGGTGTGGAAGATCGTATCGAAACCAACGAGAACAACTCCTCGTATCGAAGTT 900  
QY 301 ArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThrArgIle 320  
DB 901 CCGCGTATCAAGCAAGGCATCTGGAAGAACCCATGACGCGAGATCAGGCTACCCGATC 960  
QY 321 LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsn 340  
DB 961 AAATATGCCAGCAAGTATGCTCAAGTGCCTAACTATTGGAAGAATTTCGATCGGTATGAAC 1020  
QY 341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArgAlaPheAla 360

DB 1021 CGCGGTCTCGTCTGCTTACGTGATAGGTGCTAAGCGTCCGAGGAAGACATTCGCA 1080  
QY 361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380  
DB 1081 GACTGGATCCCTAAGAACGGCAAGAGTCTGTCTATGGCGATGTATTGCTCTCTCGAA 1140  
QY 381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400  
DB 1141 AAGCTTATAGGAAGGAGCCAGGCCAACCGCTGAGTACTTATTTTCAGCGAGACGCTC 1200  
QY 401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420  
DB 1201 TTCGGTGGTACCGAGTGGTTCGTTTGCACAGTTTGCACAGCTTGCCTTCAACGCTTGCCT 1260  
QY 421 AspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSer 440  
DB 1261 GATGCTCATGCCGGTATCCTCAATCGCTTACGACAAAGTACAAGACTACTCCTCCCTCG 1320  
QY 441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp 460  
DB 1321 CTCGACCGTAGGTGCTGCCCGCCCATGCTCCGATATTGTCGCCGCGGTATCCTCCCGAC 1380  
QY 461 LysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLys 480  
DB 1381 AAGCTCCCGCATATATTCAAGAATGTAATCGACAAGAAATTCAAAGGCGACACGAAGAAG 1440  
QY 481 TyrAlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMet 500  
DB 1441 TATGACACTTCGTATTTCGACAAGAGTGTGGTTCCTTATAGCGACAAGTTCATGCCCATG 1500  
QY 501 LeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520  
DB 1501 CTCAGTCCATGGACAGGAAAAGTTTGCACAAGCTATCGAGAAGATCCGCGCAGTAGAG 1560  
QY 521 LeuSerLysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAla 540  
DB 1561 CTTTCCAAGAGCGTAATAGCTGCTCGCGCTATTTCAGGCGCATGCGATGCCAATGCC 1620  
QY 541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560  
DB 1621 TATGCCATTGAGAAGGCGACGCTCTTTTTCGCGGTTTTCGCGTGTAGATGTACCCCGGA 1680  
QY 561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580  
DB 1681 CGTGCTCTGCCGAGGATGCCAACTTCACCATGCTATGACCTACGGCTCCATCAAGGGA 1740  
QY 581 TyrGluProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600  
DB 1741 TATGAACCGCAGGAGCGGTGCTGCTGTGTACAACTATCATACGACAGCAAGGCGGTATTGGAG 1800  
QY 601 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg 620  
DB 1801 AAGCAGATCTTAAGCGGATGAGTTGCCGTACAGGAGAATATCCTCGACCTCTTCGCG 1860  
QY 621 ThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 640  
DB 1861 ACCAAAACATATGTGCTATGCGCGAGAACGCTCAGCTCCATATCGCTTTCTTCTATCGAAC 1920  
QY 641 AsnAspIleThrGlyLysAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 660  
DB 1921 AACGACATCACGGCGGTAACTCCGCGTAGCCCGTATTTCGATAAGAAGCGCGCTGTGATC 1980  
QY 661 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAsp 680  
DB 1981 GGTCTTCTTCGATGGCAACTGGGAAGCTATGAGTGGTGACATCGAGTTCGACACCCGAT 2040  
QY 681 LeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGly 700  
DB 2041 CTGACGCGACAATCAGCGTGGACATCCGCTACGCTTCTTCTCATGATTGACAAATGGGT 2100  
QY 701 GlnCysProArgLeuIleGlnGluLeuLysLeuIle 712

Db 2101 CAGTGGCCCCGCTCTCATCCAGAGCTGAAGTTGATC 2136

## RESULT 2

US-09-815-242-9284  
; Sequence 9284, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9284  
; LENGTH: 2652  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(2652)  
; US-09-815-242-9284

Alignment Scores:  
Pred. No.: 0.00017 Length: 2652  
Score: 133.00 Matches: 106  
Percent Similarity: 33.20% Conservative: 64  
Best Local Similarity: 20.70% Mismatches: 164  
Query Match: 3.58% Indels: 178  
DB: 10 Gaps: 25

US-10-008-355-2 (1-712) x US-09-815-242-9284 (1-2652)

Qy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203  
Db 76 GATGTTTTCAAGCCTTCAGCGCATCAAAAGGCTAAGCCTTATTCAATCGTTATTCACCA 135  
Qy 204 ProSerValGlyLysPheGlyClyAspThrAspAsnTrpMetTrpProArgHisThr 223  
Db 136 CCAACGGTTACAGGTAACCTT-----CACCTT 162  
Qy 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlnTyrSer 243  
Db 163 GGT-----CACGCT 171  
Qy 244 LysAspAsnLysProTyrLysProValTyrPheAlaValSerMetGlnGlyTyrLys 263  
Db 172 TGGGATACA---ACTTTCAGATATTATCATCCGTCACAAACGCGATCGAAGTTTGTAT 228  
Qy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282  
|||||

Db 229 -----ACCTTTTGCTTCTCGGTGGTGGACACGACGAGGATTGCCACT 270  
Qy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302  
Db 271 CAGGCTAAGGTTGAGGAGCGCTTGGGTGGTGGGCGATTTCCCGCTATGACCTGGTGGT 330  
Qy 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318  
Db 331 GAGTCTTTTTCAGCAAAAGTCTGGGAATGGAAGAC-----GAATATGCCACT 378  
Qy 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337  
Db 379 ACCATCAAGGAACAATGGGCAAGATGGGCTCTCTGTAGACTATTCGTGACGGTTTC 438  
Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluArg 357  
Db 439 ACTCTTGACGAAGGTTGTCA-----AAAGCTGTTCGTAAG 474  
Qy 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373  
Db 475 GTCTTTGTGGACCTTTACAAGAAAGGCTGATCTACCGTGGTGGTATTCATCATCACTGG 534  
Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384  
Db 535 GACCCAGAGCTCGCACAGCCCTTCTGATATTGAGGTGATTCACAGGATGTGGAAGGT 594  
Qy 385 -----GluGlyAlaLysAlaAsnArgGluMetThr 394  
Db 595 GCCTTCTACCACATTAATACATGCTGGAAGATGGTTCACGGCCCTTGAAGTTGCTACA 654  
Qy 395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414  
Db 655 ACTCGTCTGAGACTATCTTTGGGGACGCTT----- 684  
Qy 415 AlaLeuAlaThrAsnProAspAlaHisIleGlyIleLeuLysSerLeuAspAspLysTyr 434  
Db 685 GCGATTGCGGTTAATCCAGAA-----GACCCGCGCTAC 717  
Qy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448  
Db 718 AAGGACTTGATTGGTTAAACGTCATCCTTCCA---ATCGCTAATAACTCATCCCAATC 774  
Qy 449 MetLeuAsp-----IleValArgArg 455  
Db 775 GTTGAGATGAGCAGCAGATCCTGAGTTTGTGTAAGTGTGCTGCTGTAATAATCACACCTGCC 834  
Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475  
Db 835 CACGATCCAAATGACTTCTTTGGTGGCCACGTCATAACTTGGCCACAAAGTCAACATCATG 894  
Qy 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495  
Db 895 AACGACGACGGAACATATCAACGAGCTTCTTTGAATTTTTCAGGCATG-----CAT 945  
Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515  
Db 946 CGTTTGAAGCTCGT-----AAGCAGCTGCTGCTTAAGTTGGAAGAAATC 990  
Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531  
Db 991 GGTGCCCTTGTCAAAATCGAAAAACGTTGCCACAGTGTGGTCACTCAGAACGTACAGGT 1050  
Qy 532 -----IleGlnAla 534  
Db 1051 GTCGTAGTTGACCCACGCTTGTCTACTCAATGGTTCTCAAGATGGACCAATTTGGCTAAG 1110  
Qy 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554  
Db 1111 AAGGCCATTGCCAACCAAGACACAGAGGACAAAG----- 1143  
Qy 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565  
Db 1144 GTCGAATTCTACCCACCTCGTTTCAACGATACCTTCTCAATGGATGGAAAAATGTCCAC 1203





```

Qy 532 -----IleGlnAla 534
Db 1051 GTCGTAGTTGAGCACCGTTGCTACTCAATGGTTCGTCAGATGGACCAATGGCTAAG 1110
Qy 535 AspAlaMetAlaAsnAlaTyAlaIleGluLysGlyLysArgLeuPheAlaGlyLeu 554
Db 1111 AACGCCATTGGCAACCAAGACACAGAGAGACAAG----- 1143
Qy 555 ArgGluMetTyProGlyArgAlaLeuProSer----- 565
Db 1144 GTCGAATTCACCCACCTCGTTTCAACGATACCTTCCATCGATGGAAATGTCAC 1203
Qy 566 AspAlaAsnPheThrMetArgMetSerTyGlySerIleLysGlyTyArgProGlnAsp 585
Db 1204 GACTGGGTATCTCGTCAGCTTGTGGGGTCAC-----CAAATC 1245
Qy 586 GlyAlaTrpTyAsnTyHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
Db 1246 CCTGCCCTGGTACAAT-----GCTGATGGTGAAATGATGTGCGCGAAGAGCTCCAGAA 1299
Qy 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
Db 1300 GGTACGGATGGATCAGGACGAGACGCTTGGAT 1335

RESULT 4
US-10-143-186-1
; Sequence 1, Application US/10143186
; Publication No. US2003007271A1
; GENERAL INFORMATION:
; APPLICANT: MENGELING, WILLIAMS L.
; APPLICANT: VORWALD, ANN
; APPLICANT: LAGER, KELLY
; APPLICANT: ROOF, MIKE
; APPLICANT: BURKHART, KELLY
; APPLICANT: GORCICA, DAVID E
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE,
; FILE REFERENCE: 27093a
; CURRENT APPLICATION NUMBER: US/10/143.186
; CURRENT FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/461.879
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 09/298,110
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 15424
; TYPE: DNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-10-143-186-1

Alignment Scores:
Pred. No.: 0.155 Length: 15424
Score: 116.50 Matches: 161
Percent Similarity: 33.12% Conservative: 100
Best Local Similarity: 20.43% Mismatches: 311
Query Match: 3.13% Indels: 217
DB: 9 Gaps: 40

US-10-008-355-2 (1-712) x US-10-143-186-1 (1-15424)

Qy 7 SerIleLeuLeuGlyAlaAlaLeuLeuGlyAlaSerGlyValAlaLysAlaAspLys 26
Db 10035 TCTAATGTTGGCGAGCAGGTTTCATCATCGCGCGCTGGT-----GCTGGGAAA 10085
Qy 27 GlyMetTrpLeuLeuAsnGluLeu-----AsnGln 36
Db 10086 ACAATAGTGGCTCCCTCAACAGGTCAGGATGGTGATGCCATTTACACCCCAACTCACCAG 10145
Qy 37 GluAsnLeuAspArgMetArgGluLeuGly-----PheThrLeuProLeuAspSer 53
Db 10146 ACCATGCTCGATATGATTAGGGCTTTTGGGACGTGCGCGGTTCACAGTCCCGAGCGGTACG 10205

```

```

Qy 54 LeuTyrSerPheAspLysProSerIleAlaAsn-----AlaValValIlePheGlyGly 71
Db 10206 ACGCTGCAATTCCTCCCTCCCGTACCGGCCCTTGGGTTCGATCCTCCTAGCGCGGT 10265
Qy 72 GlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis----- 88
Db 10266 TGGTGCTCTGGCAAGAATTCTCTCTGGATGAAGCAGCGTATTGTAATCACCTTGATGTC 10325
Qy 88 ----- 88
Db 10326 TTGAGGCTTCTTAGCAAAACTACCTCACCCTGCTGGGAGATTTCAAACAACACTCAACCA 10385
Qy 89 -----HisCysGlyTyArgIleAlaIleGlnSerGlnSer----- 99
Db 10386 GTGGGTTTGTATTCATTCATGCTATGCTTTTGCACATCATGCTCAGACTCAACTGAAGACC 10445
Qy 100 -----ThrValAspHisAspTyLeuArgAspGly 109
Db 10446 ATCTGGAGATTGGACAGAAATATCTGTGAGGCCATTTCAGGCCAGATTAC---AGGACAAA 10502
Qy 110 PheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSerValLysTyLeu 129
Db 10503 CTTGTATCCATG-----GTCAACACACACCCGTGTAACTACGTG 10541
Qy 130 ArgLysIleValLysVal-----ThrAspLysValGluGly 141
Db 10542 GAAAACCTGTCAAGTATGGCGCAAGTCTCACCCCTTACCACAGGCGGAGAGACGGC 10601
Qy 142 GlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln 161
Db 10602 -----GCCATCACAAATTGACTCCAGTCAAGGGCCACATTTGATGTGTACACTG 10652
Qy 162 GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTy 181
Db 10653 CATTTGCCACCTAAAGATTCTCAACAGCGCAAGGCCCTTGTGTCT-----ATT 10703
Qy 182 SerAsnAsnGluTyPheLeuIleValTyAspValPheLysAspValArgMetValPhe 201
Db 10704 ACCAGGCCAAGACATGTGCTCTTTGTGTATGCCACACACAGGCACTGCAGAGATGTT 10763
Qy 202 AlaProProSer-----SerValGlyLysPheGlyGlyAspThr 214
Db 10764 GATCTTCTCGGAAAGGCACACCCGTCACCTCGTGTGACCCGT-----GACGAG 10814
Qy 215 AspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyAlaGly 234
Db 10815 CAGCTGATCGTGTAGATAGAAATAACAAAGATTCACCGTTGCTCAGGCTCTAGGCAAT 10874
Qy 235 AlaAspAsnArgProAlaGluTySerLysAspAsnLysProTyLysProValTyPhe 254
Db 10875 GGGATAAATTCAGGCC-----ACAGACAAGCGCGTTGTAGATTCTCTCGCGCCC 10925
Qy 255 AlaAlaValSerMetGlnGlyTyLysAla-----AspAspTyArgAlaMetThrIleGly 272
Db 10926 ATTTGTGCAGATCTGGAAGGTGCGAGCTCCCGCTCCCAAGGTGCGCACACAACTTGGGA 10985
Qy 273 PheProGlySerThrAspArgTyLeuThrSer----- 283
Db 10986 TTTTATTCTCGCCTGAT-----TTGACACAGTTTGTCTAAACTCCCGGTAGAACTTGCA 11039
Qy 284 -----TrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArg 301
Db 11040 CCCCACTGGCCCGGT-----GTGACACCCCAACAATGAANAAG----- 11078
Qy 302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln-AlaThrArgIleLeu 321
Db 11079 -----TGCCAGACCGGTGTTGTTGAGCTTCGCCCGCTCCATAAG 11120
Qy 321 sTyAlaSerLysTyArgAlaGlnSerAlaAsnTyTrpLys----- 334
Db 11121 TATAGCCGCGGTGATCGGTCCCGCTACATGTTGGTGGGCCCTCAGTGTCTTCTGGGCACC 11180

```

```
QY 335 -----AsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleG1 350
Db 11181 CTTGGGGTTGTGTCACTACTTCACAAATTTGTCAGGGCCAGGCTCAAAATGCTTCGG 11240
QY 350 YArgLysArg-----AlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnG1 367
Db 11241 GAGACAGCTTTCAGCACCGCGCAATGAGGTAGATTGCGGTGAGTATCTCGATGACCGG 11300
QY 367 yLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
Db 11301 GAGCGA-----GAAATTCCTGAGTCCCTCCCGCATGCTTCATTTGGCGGACGT 11347
QY 387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVa 407
Db 11348 CAAAGGC-----ACTACCGTTGGAGGATGTCAACCATGT 11380
QY 407 l-----ArgPhe-----AlaGlnPheAlaAsnAlaLeuAlaTh 418
Db 11381 CACCTCCAAATACCTTCGCGGCTTCTTCCCAAGGAATCAGTCGCGGTAGTCGGGGTTTC 11440
QY 418 rAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr----- 434
Db 11441 AAGCCCGCGGAAGCGCAAGACGATTGACATTAACAGATGTATCTCCAGATCT 11500
QY 435 -LysAspTyrLeu---ProSerLeuAspArgLysValLeuProAlaMetLeuAspIleVa 453
Db 11501 CGAAGCTTACCTCCACCCAGAGACCCAGTCCAAAGTCTGGAATCATGTTGGACTTCAA 11560
QY 453 lArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLy 473
Db 11561 GGAAGTTCGACTGATGGTCTGGAAGGCAAGACGGCTTATTTCAACTTTGAAGCGCGCA 11620
QY 473 sPheLysGly---AspThrLysLysTyrAlaAspPheValPheAspLysSerValValPr 492
Db 11621 TTTCACTGGTACCAGCTTGCAGCTATGCTCGCTACATC-----CGAGTTCC 11688
QY 492 oTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAl 512
Db 11669 TGTAAAC-----TCTAC 11680
QY 512 aIleGluLysAspProAlaVal-----GluLeuSerLysSerValIleAlaAlaAlaAr 530
Db 11681 GGTGATTTGGACCCCTGCATGGCCCTTTCACAGAGAGAGTGTTCGGGTCCAC 11740
QY 530 gAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPh 550
Db 11741 TCATTGGGAGCTGACCTCGCAGTCACCCCTTAT-----GATTACGTGCGCAAAATCAT 11794
QY 550 ePheAlaGlyLeu-----ArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAs 568
Db 11795 CTTGTCTAGTCATACCATGGTGAATGCCCCCTGGGTACAAATCCTGGCGTGGCGGA 11854
QY 568 nPheThrMet-----ArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGl 586
Db 11855 GTTCTGCTTGACGATGAGTACAGGTGAGGTACAAACACCTGGGGGTTTGA--TCGGATAC 11911
QY 586 yAlaTrpTyrAsnTyrHisThrGlyLysGlyValLeuGluLysGlnAspProLysSe 606
Db 11912 ACGG---TATCTGACGAGTTCACCGGAACCGT-----GAGGACTGGGAGGATTACAA 11962
QY 606 rAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyAr 626
Db 11963 TGATCGGTTTCGTGGCGGCCAG-----AAAGGGAATTTTATAAGCG 12004
QY 626 gTyrAlaGluAsnGlyGlnLeuHisIle-----AlaPh 637
Db 12005 CACTGCCACCATGAGGTTCATTTTCCCGCGGCCCTGTTCATTTGAACCAACTTTAGG 12064
QY 637 eLeuSerAsnAsnAspIleThrGlyLysAsnSerGlySerProValPheAspLysAsnG1 657
Db 12065 CTTGAATTTGAATGAATG-----GGGTCCATGCAAGCCCTCTTTGACAAATTTGG 12115
QY 657 YArgLeuIleGlyLeuAlaPhe 664
```

```
Db 12116 CCAACTTTTCGTGGATGCTTTC 12137
RESULT 5
US-10-143-186-2
; Sequence 2, Application US/10143186
; Publication No. US20030072771A1
; GENERAL INFORMATION:
; APPLICANT: MENGELING, WILLIAMS L.
; APPLICANT: VORWALD, ANN
; APPLICANT: LAGER, KELLY
; APPLICANT: ROOF, MIKE
; APPLICANT: BURKHART, KELLY
; APPLICANT: GORCYCA, DAVID E
; TITLE OF INVENTION: PORCINE REPRODUCTIVE AND RESPIRATORY SYNDROME VACCINE,
; TITLE OF INVENTION: BASED ON ISOLATE JA-142
; FILE REFERENCE: 27093a
; CURRENT APPLICATION NUMBER: US/10/143,186
; PRIORITY FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US/09/461,879
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 09/298,110
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 15424
; TYPE: DNA
; ORGANISM: Porcine reproductive and respiratory syndrome virus
US-10-143-186-2
Alignment Scores:
Pred. No.: 0.155 Length: 15424
Score: 116.50 Matches: 161
Percent Similarity: 33.12% Conservative: 100
Best Local Similarity: 20.43% Mismatches: 311
Query Match: 3.13% Indels: 217
DB: 9 Gaps: 40
US-10-008-355-2 (1-712) x US-10-143-186-2 (1-15424)
QY 7 SerIleLeuLeuGlyAlaAlaLeuLeuLeuGlyAlaSerGlyValAlaLysAlaAspLys 26
Db 10035 TCTAATGTGTTCGCGAGCAGGTTCATCATCGCCGCCCTGGT-----GCTGGAAA 10085
QY 27 GlyMetTTrpLeuLeuAsnGluLeu-----AsnGln 36
Db 10086 ACATACTGCTCTTCAACAGGTCCAGGATGCTGATGCTATTACACGCCAACTCACCAG 10145
QY 37 GluAsnLeuAspArgMetArgGluLeuGly-----PheThrLeuProLeuAspSer 53
Db 10146 ACCATGCTCGATATGATTAGGCTTTGGGGAGCTGCCGGTTCACGTCGCCAGCAGTAGG 10205
QY 54 LeuTyrSerPheAspLysProSerIleAlaAsn-----AlaValIlePheGlyGly 71
Db 10206 AGCGTGCAATTCCTCCCTCCCGTACCGGCCCTTGGGTTCGCATCTCTAGCCGCGGT 10265
QY 72 GlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis----- 88
Db 10266 TGGTCTCTGGCAAGAATTCCTCTCGATGAAGCAGCGGTATTGTAATCACCTTGATGTC 10325
QY 88 ----- 88
Db 10325 TTGAGGCTTCTTAGCAAACTACCTCCTCTGCGGAGATTTCAAACAACACTCCACCA 10385
QY 89 -----HisCysGlyTyrGlyAlaIleGlnSerGlnSer----- 99
Db 10386 GTGGGTTTTGATTCTCTATTGCTATGTTTTTGACATCATGCTCCACAGCTCAACTGAAGACC 10445
QY 100 -----ThrValAspHisAspTyrLeuArgAspGly 109
Db 10446 ATCTGGAGATTGTGACAGAATATCTGTGATGCCATTTCACCCAGATTCAC---AGGCACAA 10502
```

QY 110 PheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSerValIleTyrLeu 129  
|||||  
Db 10503 CTGTATCCATG-----GTCAACAACACCCCGTGAACCTCAGTG 10541  
QY 130 ArgIleValIleVal-----ThrAspLysValGluGly 141  
||| |||||  
Db 10542 GAAAAACCTGTCAAGTATGGCAAGTCTCACCCCTTACCACAGGGACCGAGAGCGG 10601  
QY 142 GlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln 161  
|||||  
Db 10602 -----GCCATCAAAATTGACTCCAGTCACAGGCCACATTTGATGGTTACACTG 10652  
QY 162 GlnLeuAlaLysLysGluAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181  
||| |||||  
Db 10653 CATTTGCCCACTAAAGATTCACTCAACAGCAAGAGCCCTTGTGCT-----ATT 10703  
QY 182 SerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPhe 201  
||| |||||  
Db 10704 ACCAGGCAAGACATGTATCTTTGTGTATGACCCACACAGGCNACTGCAGAGCATGTTT 10763  
QY 202 AlaProProSer-----SerValGlyLysPheGlyGlyAspThr 214  
||| |||||  
Db 10764 GATCTCTCGGAAGGCACACCCGCTCAACCTCGCTGTGCACCGT-----GACGAG 10814  
QY 215 AspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGly 234  
||| |||||  
Db 10815 CAGCTGATCGTGTAGATAGAAATAACAAAGATGACGGTTGCTCAGGCTAGGCAAT 10874  
QY 235 AlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPhe 254  
||| |||||  
Db 10875 GGGGATAAATTCAGGGC-----ACAGACAAGCGGTGTAGATTCTCTCCGCGCC 10925  
QY 255 AlaAlaValSerMetGlnGlyTyrLysAla-----AspAspTyrAlaMetThrIleGly 272  
||| |||||  
Db 10926 ATTTGTCAGATCTGGAGGGTCAGCTCCCGCTCCCAAGGTGCGACACAACTTGGGA 10985  
QY 273 PheProGlySerThrAspArgTyrLeuThrSer----- 283  
||| |||||  
Db 10986 TTTTATTCTCGCCTGAT-----TTCACACAGTTTGTCTAAACCTCCCGGTAGAACTTGCA 11039  
QY 284 -----TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArg 301  
||| |||||  
Db 11040 CCCCACTGGCCCGTG-----GTCAACACCCAGACAATGAAAG----- 11078  
QY 302 GlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln-AlaThrArgIleLys 321  
||| |||||  
Db 11079 -----TGCCAGACCGGTGGTGTGCTACCTTCGCCCGCTCCATAAG 11120  
QY 321 sTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLys----- 334  
||| |||||  
Db 11121 TATAGCCGCGGTGCATCGGTGCGGGGTACATGTTGGGCCCTCAGTGTCTTCTGGGCACC 11180  
QY 335 -----AsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGln 350  
||| |||||  
Db 11181 CTGGGGTTGTGTACTACTATCTCACAAAATTTCTCAGGGCGGAGGCTCAAAATGCTTCCG 11240  
QY 350 yArgLysArg-----AlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGln 367  
||| |||||  
Db 11241 GAGACAGTCTTCAGCACCGCGCGAATTGAGGTAGATTGCCGTGAGTATCTTGATGACCGG 11300  
QY 367 yLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387  
||| |||||  
Db 11301 GAGCGA-----GAAATTGCTGAGTCTCCCTCCCATGCTTTTCATTGGCGAGCT 11347  
QY 387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVal 407  
||| |||||  
Db 11348 CAAAGGC-----ACTACCGTTTGGAGGATGCTCCACATGT 11380  
QY 407 l-----ArgPhe-----AlaGlnPheAlaAsnAlaLeuAlaLys 418  
||| |||||  
Db 11381 CACCTCCAAATACCTTCCGCGCTCTCTCCCAAGGAATCAGTCGCGGTAGTCGGGGTTTC 11440  
QY 418 rAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr----- 434

Db 11441 AAGCCCGGGAAAGCCGCAAAAGCAGTTTGCACATTAAACAGATGTGTATCTCCAGATCT 11500  
QY 435 -LysAspTyrLeu---ProSerLeuAspArgLysValLeuProAlaMetLeuAlaLeu 453  
||| |||||  
Db 11501 CGAAGCTTACTTCCACCAGACAGCCAGTCCAGTCTGGAATGATGTGTGACCTCAA 11560  
QY 453 lArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLys 473  
||| |||||  
Db 11561 GGAAGTTTCGACTGATGTCTGGAAGGACAAGCGGCTATTTTCAACTTGAAGCCGCCA 11620  
QY 473 sPheLysGly---AspThrLysLysTyrAlaAspPheValPheAspLysSerValValPr 492  
||| |||||  
Db 11621 TTTCACTGTGTACAGCTTGTCAAGCTATGCTCGTACATC-----CGAGTTTC 11668  
QY 492 oTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAl 512  
||| |||||  
Db 11669 TGTTAAC-----TCTAC 11680  
QY 512 aileGluLysAspProAlaVal-----GluLeuSerLysSerValIleAlaAlaAlaAr 530  
||| |||||  
Db 11681 GGTGTATTTGGACCCCTGCATGGCCCTTTCGCAACAGAGAGTGTGTGGGTCCAC 11740  
QY 530 gAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPh 550  
||| |||||  
Db 11741 TCATTGGGGAGCTGACCTCGCAGTCAACCCCTTAT-----GATTACGTCGCCAAATCAT 11794  
QY 550 ePheAlaGlyLeu-----ArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAs 568  
||| |||||  
Db 11795 CTGTCTAGTGCATACCATGTTGAAATGCCCTTGGGTACAAATCTGGCGTCCGCGGA 11854  
QY 568 nPheThrMet-----ArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGln 586  
||| |||||  
Db 11855 GTTCTCGCTTGACGATCCAGTGAGGTACAAACACACCTGGGGTTTGA-----TCGGATAC 11911  
QY 586 yAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSe 606  
||| |||||  
Db 11912 AGCG---TATCTGTACGAGTTTCAACGGAAACGGT-----GAGGACTGGGAGGATTACAA 11962  
QY 606 rAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyAr 626  
||| |||||  
Db 11963 TGATCGGTTCTGTCGGCGCCAG-----AAGGGAATAATTTATAAGGC 12004  
QY 626 gTyrAlaGluAsnGlyGlnLeuHisIle-----AlaPh 637  
||| |||||  
Db 12005 CACTGCCACACGATGAGTTTTCATTTTCCCGCGGCCCTGTCTTGAACCACTTTAGG 12064  
QY 637 eLeuSerAsnAsnAspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGln 657  
||| |||||  
Db 12065 CCTGAATGAAATGAAATG-----GGTCCATGCAAGCCCTCTTTGACAAATTTGG 12115  
QY 657 yArgLeuIleGlyLeuAlaPhe 664  
||| |||||  
Db 12116 CCAACTTTTCTGGATGCTTTC 12137  
RESULT 6  
US-09-925-301-170  
; Sequence 170, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US/09/925,301  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 170  
; LENGTH: 3598

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (16)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (22)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (964)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-170

Alignment Scores:
Pred. No.: 0.0203 Length: 3598
Score: 116.00 Matches: 110
Percent Similarity: 33.80% Conservative: 85
Best Local Similarity: 19.06% Mismatches: 198
Query Match: 3.12% Indels: 185
DB: 10 Gaps: 30

US-10-008-355-2 (1-712) x US-09-925-301-170 (1-3598)

Qy 90 CysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGly 109
Db |||||:|||||
39 TGTGTTTCGGTGAG-----CTCTCAGACCGA 65

Qy 110 PheValSerArgThrMetGlyGluGluLeu-----Pro 120
Db |||||:|||||
66 TTTCAGCGTCGTCGGGACAGCGTGCAGAGTCGRCTCCTCAGACATGGCGGCT 125

Qy 121 IleProGlyLeuSerValTyrLeuArgIleValLysValThrAspLysValGlu 140
Db |||||:|||||
126 CCACGCGC-GCGGGGAAGCAGCGGATCCCAAGTGGCCAAAGTGAAGAAAGAGAA 184

Qy 141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCys 160
Db |||||:|||||
185 CCGTAGGTACAG--ATACTGCTGAA--CAACTCTTAAGAGAGCGCTAAGAAAGAGAA 238

Qy 161 GlnGluLeu-----AlaLysLysGluAsnAlaAspGluAsnGlnLeuCystile 176
Db |||||:|||||
239 CTGTAGCTTCTCCACCTCCACCTCAACAGAGATCAGATGAAGAAATTA----- 292

Qy 177 ValGluProPheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAsp 196
Db |||||:|||||
293 -----AATGATTATAACTTAAGGAAAGGAAAGAACTTTTGAAGAT 331

Qy 197 ValArgMetValPheAlaProSerValGlyLysPheGlyLysProVal 216
Db |||||:|||||
332 -----AATATAAGAAAGAAAGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 364

Qy 217 TrpMet-----TrpProArgHisThrGlyAspPheSerValPheArgValTyr 232
Db |||||:|||||
365 TGGATAAAATACGCACAAATGG----- 385

Qy 233 AlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProVal 252
Db |||||:|||||
386 -----GAAGAAAGCGCTAAAGGAGATTCMAAGGCGCTCGATCCATA 424

Qy 253 TyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGly 272
Db |||||:|||||
425 TACGAGCGTGTCTTAGATGTA-----GACTAC----- 451

Qy 273 PheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGluAsn 292
Db |||||:|||||
452 -----CGAAATATTACCTCTGGCTGAAATACGACGAGAAATGGAATG 493

Qy 293 GluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMet 312
Db |||||:|||||

```

```

Db 494 AAGAAATCCCAAGTCAMCCATGCTCGA-----AATATCTGGACCGGCCATA 541
Qy 313 SerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332
Db |||||:|||||
542 -----ACAACGCTGCTCGAGTTAATAGTTCTGGTACAAAGTACACGTAC 586
Qy 333 TrpTysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLys 352
Db |||||:|||||
587 ATGGAGGAAATGTTGGGA-----AAGGTTGCCGGTCCCGG 622
Qy 353 ArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyr 372
Db |||||:|||||
623 CAGGTGTTTGAGCGC-----TGATAGAGTGGCAGCTGAGGAGCAACGCTGG 670
Qy 373 GlyAspValLeuSerSerLeuGlyLysAlaTyrLysGlyAlaLysAlaAsnArgGlu 392
Db |||||:|||||
671 CACTCTACATC--AACTTTGAGCTGAGATACAAGAG-----GTGATCGGGCC 718
Qy 393 MetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluVal-----ValArgPhe 409
Db |||||:|||||
719 CGCACCATTTATGACCGAKTGTCTCTCGTCGACCCCTGATTTAAGAACTGGATCAAGTAT 778
Qy 410 AlaGlnPheAlaAsnAlaLeuAla----- 417
Db |||||:|||||
779 GCCCGCTTTGAAGAAAAACATGCTTATTTTGGCCATGCACGGAAGAGTATGAGAGAGCT 838
Qy 418 -----ThrAsnProAspAlaHisAlaGlyIle--LeuLysSerLeu 430
Db |||||:|||||
839 GTGGAATCTTGTGGAGATGAACATATGATGATGACACCTTATGTTGCTTGGCCAAAGTTT 898
Qy 431 AspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeu 450
Db |||||:|||||
899 GAAGAAATTCAGAAAGAGTTT-----CAAGGGTACGAGTATTTACAAGTATGCCCTG 952
Qy 451 AspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsn----- 468
Db |||||:|||||
953 GAC-----AGAAATTTNAAACAGATGCCCAAGAACTCTTTAAAAATTTATACC 1000
Qy 469 ValIleAspLysLysPheLysAspThrLysLysTyrAlaAspPheValPheAspLys 488
Db |||||:|||||
1001 ATCTTTGAGAGAAAGTTT---GGTGATAGCGGGGTATTTGAAGATATCATTTGTGAGCAAA 1057
Qy 489 SerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLys---Glu 507
Db |||||:|||||
1058 CGGAGATTCAGTACGARGAAGAGAGTGAAGCGAATCCACACAATATGATGATGCTGTTT 1117
Qy 508 LysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSerValIleAla 527
Db |||||:|||||
1118 GATTACTTTCGCTTGGTAGAAGTGACGCAAGCTGAA----- 1156
Qy 528 AlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLys 547
Db |||||:|||||
1157 GCCGTGAGAGAGTCTATGAAGGCGCAATGCCAATGCCACCATTCCAGGAGAGAGG 1216
Qy 548 ArgLeuPheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAla 567
Db |||||:|||||
1217 CAC-----TGAAGCGCTACATTTAT----- 1237
Qy 568 AsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAla 587
Db |||||:|||||
1238 -----CTT 1240
Qy 588 TrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAsp 607
Db |||||:|||||
1241 TGGATCAACTATGCACTCTATGAAGAA-----TTGAGGAGCAAGGATCCTGAGAGGACA 1294
Qy 608 GluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyr 624
Db |||||:|||||
1295 AGACAGGTGTATCAAGCCTCT---TTGGAACATAATTCCTCACAAAAAGCTC 1342

```

RESULT 7  
 US-09-815-242-6368  
 ; Sequence 6368, Application US/09815242

Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR FILING DATE: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6368  
LENGTH: 2856  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(2856)  
US-09-815-242-6368

## Alignment Scores:

Pred. No.:	0.0272	Length:	2856
Score:	113.50	Matches:	118
Percent Similarity:	30.56%	Conservative:	69
Best Local Similarity:	19.28%	Mismatches:	144
Query Match:	3.05%	Indels:	282
DB:	10	Gaps:	30

US-10-008-355-2 (1-712) x US-09-815-242-6368 (1-2856)

Qy	150	MetGluArgLeuArgLysAlaGlnGluValCysGlnGluLeuAlaLys-----	165
Db	1	ATGGAAAAGACATATAACCCACAAGATATCGAACACGCGCTTTACGACGCACTGGGAAAAG	60
Qy	166	-----LysGluAsnAlaAspGluAsnGln-----LeuCysIleValGluPro	179
Db	61	CAGGGCTACTTTAAAGCCATAATGGCGATGAAAGCCAGGAAAGTTTCTGCATCATGATCCCG	120
Qy	180	PheTyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMet	199
Db	120	-----	120
Qy	200	ValPheAlaProProSerSerValGlyLysPheGlyAspThrAspAsnTrpMetTrp	219
Db	121	-----CCGCCG-----	126
Qy	220	ProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgPro	239
Db	127	-----AACGTCAC-CGGCAGTTTGGCATATGGTTCACGCGCTTCCA	164
Qy	240	AlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMet	259
Db	165	GCAACCATCATGGATACCATGATCCG-----CTA	194

Qy	260	GlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArg	279
Db	195	TCAGCGCATGCAAGGCAAAAACACCCCTGTGCAGGTGGTACTGACACGCGGGATCGC	254
Qy	280	TyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnPro---Argile	298
Db	255	TACCCAGATGGTGTGAGCGCAAGATTGCCGAGAGAAGTAACCCCTCAGACTA	314
Qy	299	GluValArg-----GlyIleLysGlnGlyIleTrpLysGluAla	311
Db	315	CGGCGCGAAGCTTTCATCGACAAANATCTGGAAATGGNAACGGANATCTGGCGGCACCAT	374
Qy	312	MetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGln-SerAlaAs	331
Db	375	TACCGTCAGATGCGCGCTCGG-----CAACTCCGTGCA	410
Qy	331	nTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyAr	351
Db	411	CTGGAGCGGTGAACGCTTTCACCATGGAGAGGCGCTGTCCAAT-----GCGGT	458
Qy	351	GlySargAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaVa	371
Db	459	CAAGAACTTTTCGTCTGTATTAAGAAGACCTGATTTACCGTGCCAAACGCTCGT	518
Qy	371	IleTyrGlyAsp-----ValLeuSerSerLeuGluLysAlaTyrLysGlu--	385
Db	519	AACTGGGATCCGAAAGTGCACCGCTATCTCTACCTGGAAGTCGAAACCGCGAATC	578
Qy	386	-----GlyAlaLysAlaAsnArgG1	392
Db	579	GAAAGGTTTCGATGTGGCACATCCGCTATCCGCTGCTGACGTCGCGAAAACCGCAGACGG	638
Qy	392	uMetThrTyrLeu-----SerGluThrLeuPheGlyGlyThrGluVa	406
Db	639	TAAAGATTATCTGGTGGTCGCGACTACCCCTCCAGAAACCTCTGTCGCGTACTGGCGT	698
Qy	406	IValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyI1	426
Db	699	A-----GCCGTTAAACCCGGA-----	714
Qy	426	eLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValle	446
Db	715	-----GATCGCGCTTACAAAGATCTGATGGCAATAT-----	747
Qy	446	uProAlaMetLeuAspIleValArgArgIlePro-----AlaAspLysLeuProAs	464
Db	748	---GTCATTCTGCGGTGGTTAACCGTCGTATTCGGATCGTTCGCGACGACACCGCGA	803
Qy	464	pIlePheLysAsnVal-----	469
Db	804	CATGGAAGAGGCAACCGGCTGCGTGAAATCACTCCGCGCACGACTTTTAACGACTATGA	863
Qy	470	IleAspLysLys-----PheLysGlyAspThrLyl	479
Db	864	AGTGGGTAAGCGTCACGCCCTGCCGATGATCAACATCCTGACCTTTGACGCGATATCCG	923
Qy	479	sLysTyrAlaAspPheValPheAsp-----LysSerValValProTyrSerAs	495
Db	924	TGAAGCCGCCAG---GTGTTCCGATACCAAGGTAAACAACTGACGTT---TATTCAG	977
Qy	495	PLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLy	515
Db	978	CGAATCTCTGCAGATGCCAGAAACTG-----GAGCGTTTGTCT-----	1017
Qy	515	sAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla-	531
Db	1018	-----GCACGTAAGACAGTCGTTCGCCGAGTTGACGCGCTTGGCGCTGCT	1061
Qy	531	-----	531
Db	1062	GGAAGAAATTAACCGCACACCTGACCGTTCCTTTACGGCGACCGTGGCGGCTAGTTAT	1121

```
Oy 532 -----IleGlnAlaAspAlaMetAlaAsn-----Al 540
Db 1122 CGAACCAATGCTGACCGACAGTGTACGTGCTGCCGATGTCCTGGCGAAACCGCGGT 1181
Oy 540 aTyraLalleGluLysGly-----LysArgLeuPhePh 551
Db 1182 TGAACGGTGTGAGAACGCGGACATTCAGTTCCGTACCGAGCAGTACGAAACATGTACTT 1241
Oy 551 eAlaGlyLeu-----ArgGluMetTyrProGlyArgAl 562
Db 1242 CTCCTGGATGCCGATATTCAGGACTGGTGTATCTCTCGTCGTTGTGTGGGGTCACCG 1301
Oy 562 aLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrCl 582
Db 1302 TATCCCG----- 1308
Oy 582 uProGlnaspGlyAlaTptTyrAsn----- 590
Db 1309 -----GCATGGTATGACGAGCGGTAACGTTTATGTTGGCCGCAACGAAGA 1355
Oy 591 -----TyrHisThrThrGlyGlyValLeuGluLysGlnAspProLysSe 606
Db 1356 CGAAGTGGTGAAGAAATAACCTCGTGTGTGATGTTGCTCGCTCAGGAC----- 1407
Oy 606 raspGluPheAlaValGlnGluAsnIleLeuAsp 617
Db 1408 -----GAAGACGTTCTCGAT 1422
RESULT 8
US-08-781-986A-221
; Sequence 221, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781.986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 221:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10758 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-221
Alignment Scores:
Pred. No.: 0.251 Length: 10758
Score: 112.50 Matches: 100
```

```
Percent Similarity: 35.39% Conservative: 78
Best Local Similarity: 19.88% Mismatches: 168
Query Match: 3.03% Indels: 158
Db: 7 Gaps: 21
US-10-008-355-2 (1-712) x US-08-781-986A-221 (1-10758)
Oy 246 AsnLysProTyrLysProValTyrPheAlaAlaVal---SerMetGlnGly----- 261
Db 341 AACAGGACTACGAGGTGCTTATATCCAAAGTGTAGAAAATATCCAAAGGTGCGAAAGAT 400
Oy 262 -----TyrLysAlaAspTyrAlaMet 269
Db 401 TATCATGAAGGAAAACTGATCATATTACCGGATTGAAGAAAAATAGACTACTACTATG 460
Oy 270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrp----- 284
Db 461 CAGGTTACATTGTGATAAAAAACAAGAAATATTACTTTAACAGGATTTTATTACTGGACCTTTA 520
Oy 285 -----GlyValGluAspArgIleGluAsnGluAsnAsn 295
Db 521 TTAAGTAAAAAATATTTATCAGATGTACCAATTTAAGATTTAGCGAAATCAGATAAAATC 580
Oy 296 ProArgIleGluValArgGlyIleLysGlnGlyIleTyrLysGluAlaMetSerAlaAsp 315
Db 581 CGAAAATATCTCTATT---GGTATTGGACCGTATAAAGTTAAGAAAAATCGTCCAGGTGAG 637
Oy 316 GlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsn 335
Db 638 GCTGTTCAACTCGTTAAATTT-----GATGATTATTGG----- 670
Oy 336 SerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 355
Db 671 -----CAAGTAAGCCTGCACCTAGACAAAATCAATTTAAAGATTATTGAT 715
Oy 356 GluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspVal 375
Db 716 CAA-----CGCGAAATTT 727
Oy 376 LeuSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyr 395
Db 728 ATTAAGGCAATGGAAGAAA----- 745
Oy 396 LeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAla 415
Db 746 -----GGCGATATTGATGTTGGCATGATGCT-----ACCGGTGCA 781
Oy 416 LeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLys 435
Db 782 ATGCGAAAAGATGCTAAGTCATCTAATGCTGCTCAAGGTATTATCT----- 829
Oy 436 AspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArg 455
Db 830 -----GCGCCCAAGCTTAGACTACGGTTTAAATAGGTTCTGTCATCATGATACGATAAAA 883
Oy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
Db 884 AAGCTA-----ATAAACTGGTAAAGTCAGACCAAAATATGAA-GACAAAGAAATTACGT 936
Oy 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db 936 ----- 936
Oy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
Db 937 -----AAGCAATGCTTTTATGCAATTTAGAGAAAATGATCAATCAAGCGTTTTCAT 990
Oy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAla--- 534
Db 991 GGTACGCTAGTGAATCAAT---AGTTTGTACCATCTCATGTCATGTCAGCACCAAT 1047
Oy 535 -----AspAlaMetAlaAsnAlaTyrAlaIleGlyLysGlyLysArgLeuPhePheAla 552
```

	:::    :::	
Db	238 GAAATATGTAATAAATAAGTGGTGAGAGCTATGCAAAATCAACTAAAAAGCGACATACA	297
QY	142 GLnLeuLysGlyIleThrAspGluMetGluArgLeuAlaGlnGlyValCysGln	161
Db	298 ATTACTGTAGCTCTAGTTAACGAGTTGAACACATTAG-----AACGAGTATTGTGAAT	351
QY	162 GluLeuAlaLysGlySerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArg	181
Db	352 AAAATAGTTGAATCAACCTCAGAAAGCCAACTACAGATGATCATGATGATGAGAGTCGATCA	411
QY	182 SerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPhe	201
Db	412 AAAGTAGATGAA-----	423
QY	202 AlaProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArg	221
Db	424 -----CGTGTGCTGAAGTTTGAAGAAGGACTCATCTTCT-----TCG	459
QY	222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgPro-----	239
Db	460 TCAAGTTCAGACTCTTCC-----ACTAAACCGGAAGCT	492
QY	240 AlGluTyrSerLysAspAsnLysProTyrLysPro-----	251
Db	493 TCAGATACAGCGAAGCCAACAGCGCAGACAAACCAGGAGAAAAGGTAGCAGAAAGCTAAG	552
QY	252 ----ValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMet	269
Db	553 AAGAGGTTGAAGAGCTGAGAAAAGCCAGGATCAAAGAAGAGAGATCGTCGT---	609
QY	270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTyrGlyValGluAspArg	289
Db	610 -----AACTACCCACCATTACTTACAAAACGCTTGAACTTGAATTCCTGAGTCCGAT	663
QY	290 IleGlu-----	298
Db	664 GTGGAAGTTAAAAAGCGGAGCTTGAACTAGTAAAGGTGAAGAGCTTAACGAACCTCGAGAC	723
QY	299 GluValArgGlyIleLysGlnGlyIleTyrLysGluAlaMetSerAlaAspGlnAlaThr	318
Db	724 GAGCAAAA--ATTAGCAAGCAGAAAGCGGAGGTTGAGAGTAAACNAGCTGAGGCTACA	780
QY	319 ArgIleLys-----TyrAlaSerLysTyrAlaGlnSerAlaAsn	331
Db	781 AGCTTAAAAAATCAAGACAGATCTGTAAGAGCAGAGAAGAGAGCTTAACGAAGACCA	840
QY	332 TyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg	351
Db	841 GATGCTAAAGAGCAA--GGTAAACCAAAGGGCGGCCAAAACGAGGAGTTCTCTGGAGAC	897
QY	352 LysArgAlaGluGluArgAlaPheAlaAspTyrPheArgLysAsnGlyLysSerAlaVal	371
Db	898 CTAGCAACACCTGATAAAAAGAAAATGAT-----GCGAAGCTCTTCAGATTACGCTGA	951
QY	372 TyrGlyAspValLeuSerSerLeuGlu---LysAlaTyrLysGluGlyAlaLysAlaAsn	390
Db	952 GTGAAGAAACTCTTCAAGCCCATCCCTGAAACCCAGAAAAAGAGGTAGCAGAACTCGAG	1011
QY	391 ArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAla	410
Db	1012 AAGAAGGTTCAAGAAGCTAAGAAAAAAGCCGAGGATCAAAAAAGAGAGATCGCGGTAAAC	1071
QY	411 GlnPheAlaAsnAlaLeuAlaThr-----AsnProAspAla	422
Db	1072 TACCCAAACCAACTACTTACAAAACGCTTGAACCTTGAATTTGCTGAGTCGCGATGTGAAGTT	1131
QY	423 HisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAsp	442
Db	1132 AAAAAACCGGAGCTTGAAGCTAGTAAAGAGGAGCTAAGCAA-----CCTCGAAACGAG	1189
QY	443 ArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAspLysLeu	462









TELEFAX: 858/677-1465

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 2319 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 1...2316

SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-10-121-032-14

Alignment Scores:

Pred. No.:	0.0486	Length:	2319
Score:	110.00	Matches:	155
Percent Similarity:	33.13%	Conservative:	105
Best Local Similarity:	19.75%	Mismatches:	268
Query Match:	2.96%	Indels:	258
DB:	9	Gaps:	42

US-10-008-355-2 (1-712) x US-10-121-032-14 (1-2319)

```

Qy 74 ThrGly---IleThrValSerAspGlnGlyLeuIlePheThrAsnHisCysGlyTyr 92
Db 174 ACAGGCAAGGTGCAACAAGGTGATCGAGGC-----TTTCTGACCAATCCTGTGGATAC 227

Qy 93 GlyAlaIleGlnSer---GlnSer-ThrValAspHisAspTyrLeuArgSpGlyPheVal 111
Db 228 GAAAAAGAAAGAACTCTTCAAGGTACTCTGTCAC----- 261

Qy 111 lSerArgThrMetGlyGluLeuProIleProGlyLeuSer----- 125
Db 262 -----GHAAGAAAGATCCCTCAGAGTGGAAAGGCCGATCCACCGGA 308

Qy 126 -----ValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyG1 142
Db 309 CATAGACGTGACGAACTACGTGAGAATCGTCTTCTGAAATCCCTGAAAGAAAGACCT 368

Qy 142 nLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnG1 162
Db 369 CAGAAAGACGTGGAACTCATAGAGAGGTTACAAACCGGCAAGAGTCAATCATGATGA 428

Qy 162 uLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyrSe 182
Db 429 G-----ATCCTGGACGACTACTATTA 449

Qy 182 rAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPheAl 202
Db 450 CGATGGAGAG---CTGGAGCGCTATATCTCCAGAGAGACGATATTCAGAGTCTGGTC 506

Qy 202 aProProSerSer-----ValGlyLysPheGlyGlyAspThrAsp----- 215
Db 507 CCCCGTTCTAAGTGGTAAAGTGCTCTCTCTCAAAAACGGAGAAGACACAGAACCCTGA 566

Qy 216 -----AsnTrpMetTrpProArgHisThr---GlyAs 225
Db 567 CCAGGTGTGAACATGGAATACAGGGAACCGGGTCTGGGAAGCGGTGTTGAAGCGCA 626

Qy 225 pPhe---SerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerly 244
Db 627 TCTCGACGGAGTGTCTACCTCTAT-----CAGCTGGAAAACTACGGAAA 671

Qy 244 sAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAl 264
Db 672 GATCAGAACACCGTCGATCCCT---TATTTCGAAAGCGGTT-----TAGCGAAA 716

Qy 264 aAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTr 284
Db 717 CAACCAAGAGACCGCGGTGTGAATCTGCCAGGACAAACCCAGAA-----GGATG 767

Qy 284 pGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGlyIlely 304

```

```

Db 768 G-----GAAACACACAGGGGACCGCAAAATCGAA-----GGATACGA 803

Qy 304 sGlnGlyIleTrpLysGluAlaMetSerAlaAsp-----GlnAlaThrArgI1 320
Db 804 AGACGCGATAATCTATGAAATACACATACGCGACATCACAGGACTCGAAAATCCTCCGGGT 863

Qy 320 eLysTyrAlaSerLysTyr---AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMe 339
Db 864 AAAAAACAAGGCCTCTATCTCGGCTCACCGAAGAAACACAGAAAGACCGGGCGGTGT 923

Qy 339 tAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPh 359
Db 924 GACAACAGGCCTTTCGCACCTTGTGGAACCTGGTGTACACACCTTCATATCTTCTT 983

Qy 359 eAlaAspTrp-----IleArgLysAsnGlyLysSerAlaValTyr----- 372
Db 984 CTTTGATTTCTACACAGCGGACGAACTCGATAAAGATTT-CGAGAAGTACTACAACCTGGG 1042

Qy 373 -----GlyAspValLeuSerSerLeuGluLysAl 382
Db 1043 GTTACGATCCTTACCTGTTCATGTTCCGGAGGCGAGTACTCAACCGATCCCAAAAC 1102

Qy 382 aTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheG1 402
Db 1103 CACACACGAGAAATCAGAGAAGTCAAAGAATGGTCAAAAGCCCTTCACAACACGGTATAG 1162

Qy 402 y-----GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAs 419
Db 1163 GTGTGATTATGGACATGTTCCCTTCACACCTACGATATAGCGCAACTCTCTCGCTTCG 1222

Qy 419 nProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLys----- 433
Db 1223 ATCAGACGGTGCCTACTACTTCTACAGAATCGACAAGACAGGTGCCTATTTCAGACGAA 1282

Qy 434 -----TyrLysAspTyrLeuProSerLeuAs 442
Db 1283 GCGGATGTGTAACTGTCATCGCAAGCGAAGACCCATGATGAGAAAATTCATAGTCGATA 1342

Qy 442 pArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAla-AspLysL 462
Db 1343 CGTCACCTACTGGGTAA-----GGAGTATCACAATAGACGGATTCAGGTTTCGATCAGA 1396

Qy 462 euProAspIlePheLysAsnValIleAspLysPheLysGlyAspThrLysLysTyr- 481
Db 1397 TG-----GGTCTCATCGACAAAAGACAATGCTCGAAGTCGAAAGACTC 1441

Qy 482 -----AlaA 483
Db 1442 TTCATAAAATCGATCCAATCATCTCTACGCGCAACCGTGGGTGGATGGGAGCAC 1501

Qy 483 spPheValPheAspLysSer-----ValValProTyrSerAspLysP 497
Db 1502 CGATCAGGTTTGGAAAGAGCGATGTGCGCGGCACACACGTGCGCAGCTTTTCAACGATGAT 1561

Qy 497 heHisAlaMetLeuLys-----S 503
Db 1562 TCAGAGACCATATAGGGTTCCTGTTCAACCCGAGCGCTCAAGGGATTCGTTCATGGGAG 1621

Qy 503 erMetAspLysGlu-----L 508
Db 1622 GATACGGAAGGAAACCAAGATCAAAAGGGTGTGTTGGAAGCATATAACTACGACGGAA 1681

Qy 508 ysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer----- 522
Db 1682 AACTCATCAAAAGTTTCGCCCTTTCGCTCCAGAGAAGAACTATAAATACGCGCGTGTCCAG 1741

Qy 523 -----LysSerValIleAlaAla-----AlaArgAlaI 532
Db 1742 ACAACCAACACTGTGGGCAAGAACTACCTTCCGCCCCCAAAAGCTGATAGAAGAAAGAA 1801

Qy 532 leGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheA 552

```



```
Qy 373 -----GlyAspValLeuSerSerLeuGluLysAl 382
Db 1043 GTTAGCATCTTACCTGTTTCATGTTCCGGAGGCGAGACTCAACCCATCCCAAAACC 1102
Qy 382 atYrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheG1 402
Db 1103 CACACACGAGAAATCAGAGAAGTCAAAAGAAATGGTCAAAAGCCCTTCACAAACACGGTATAG 1162
Qy 402 y-----GlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAs 419
Db 1163 GTGTGATTTATGGACATGGTGTTCCTCCATCACCTAGCTAGGTAGCGGAATCTCTCGGTCCG 1222
Qy 419 nProAspAlaHisAlaGlyIleLeuLysSerLeuAspLys----- 433
Db 1223 ATCAGACGGTCCGCTACTACTTCTACAGAATCGACACACAGGTCCTATTTTGAACGAAA 1282
Qy 434 -----TyrLysAspTyrLeuProSerLeuAs 442
Db 1283 CCGGATGTGTAACGTCATCGCAAGCGAAAGACCCTATGATGAGAAAATTCATAGTCGATA 1342
Qy 442 pArgLysValLeuProAlaMetLeuAspIleValArgArgGluIleProAla-AspLysL 462
Db 1343 CCGTCACTACTGGGTAAA-----GGAGTATCATATAGACGGATTTCAGGTCGATCAGA 1396
Qy 462 euProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyr- 481
Db 1397 TG-----GGTCTCATCGACAAAAGACAAATGCTCGAAGTCGAAAGAGCTC 1441
Qy 482 -----AlaA 483
Db 1442 TTCATAAATCGATCCAACTATCATTTCTACGGCGAAACCGTGGGTGGATGGGAGAC 1501
Qy 483 spPheValPheAspLysSer-----ValValProTyrSerAspLysP 497
Db 1502 CGATCAGGTTTGGAAAGACCGATGTCGCCGCCACACACCTGGCAGCTTTCAACGATGAGT 1561
Qy 497 heHisAlaMetLeuLys-----S 503
Db 1562 TCAGAGACCAATAAGGGTTCCGGTTTCAACCCGCGGTCAGGGATTTCGTATGGGAG 1621
Qy 503 erMetAspLysGlu-----L 508
Db 1622 GATACGGAAAGGAACCAACCAAGATCAAAAGGGGTGTTGTTGGAAGCATAAACTACGACGGAA 1681
Qy 508 ysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer----- 522
Db 1682 AACTCATCAAAAGTTCCCTCCCTTGATCCAGAGAAGCACTATAAACTACGACGGTGTACG 1741
Qy 523 -----LysSerValIleAlaAla-----AlaArgAlaI 532
Db 1742 ACAACACACACTGTGGGACAAGAACTACCTTCCGCCCAAAAGCTGATAAGAAAAGGAAT 1801
Qy 532 leGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheA 552
Db 1802 GGACCGAAGAAGAACTGAAAACGCCCAAGAACTGGCTGGTGGCGATACTTCTCACTTCTC 1861
Qy 552 laGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMeta 572
Db 1862 AAGGTGTTCTTCTCCACGGAGGGGAGGACTCTCGACGACGACGAGCAATTTTCAACGACA 1921
Qy 572 rgMetSerTyrGly-----SerIleLysGlyTyrGluProGlnAspGlyAlaTrp- 588
Db 1922 AC---TCCTACAACGCCCTATCTCGATAAAGCGGTTCGATTACGAAAGAAAATTCAGT 1978
Qy 589 -----TyrAsnTyrHisThrThrGlyLysGlyVal-----LeuGluLysGlnA 603
Db 1979 TCATAGACGTGTTCAATTTACCAC-----AAGGGTCTCATAAACTCAGAAAAGAAC 2029
Qy 603 spProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLysA 623
Db 2030 ACCCT-----GCTTTCAGGCTGAAAA 2050
Qy 623 snTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn----- 640
```

```
Db 2051 AC-----GCTCAAGAGATCAAAAAACACCTGGAATTTCTCCCGGGCGGGAGAA 2098
Qy 641 -----AsnAspIleThrGlyGlyAsnSerGlySerPro-----ValP 653
Db 2099 GAATAGTTGCGTTTCATGCTTAAGACCACGCGAGTGGTGGATCCCTCGGAAAGACATCGTGG 2158
Qy 653 heAspLysAsnGlyArgLeuIleGlyLeuAlaPhe-----AspGlyAsnTyr----- 668
Db 2159 TGATTTACAATGGAAACTTAGAGAAGACACATACAAACTGCCAGAAAGAAAATGGAATG 2218
Qy 669 -----GluAlaMetSerGlyAspIleG 676
Db 2219 TGGTTGTCAACAGCCAGAAAGCCGGAACAGAAGTGATAGAAACCTGCAAGGAACAATAG 2278
Qy 676 luPheGluPro 679
Db 2279 AACTCGATCCG 2289

RESULT 13
US-10-027-806-61
; Sequence 61, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOIP-002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 61
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2538)
US-10-027-806-61

Alignment Scores:
Pred. No.: 0.0555 Length: 2538
Score: 110.00 Matches: 121
Percent Similarity: 37.06% Conservative: 88
Best Local Similarity: 21.45% Mismatches: 233
Query Match: 2.96% Indels: 123
DB: Gaps: 26

US-10-008-355-2 (1-712) x US-10-027-806-61 (1-2538)
Qy 20 GlyValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlu-----LeuAsn 35
Db 901 GCGTGGCGGATTCGACATACCCCTGTATGATGCGGGATTCGCCACGCTCGGCAC 960
Qy 36 GlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyr 55
Db 961 GCGGTCCATCTGGACCTGTACAGACC-----TTCTCGAACAGGTCTTCAGCTGTAT 1014
Qy 56 SerPhe-----AspLysProSerIleAlaAsnAlaValIle 68
Db 1015 GCATTGGCGGAAAGTATACAGATTACTCCCTGAACACGCTGTCCAAAGCGATGCTCGGC 1074
Qy 69 PheGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
Db 1075 GAGGCAAGGTCGATTATGGCTGTCTCTCGGGGATCTCACTCTATACCAGACTGCAAC 1134
Qy 89 HisCysGlyTyrGlyAla-----IleGlnSerGlnSerThrValAspHisAspTyrLeu 106
Db 1135 TATTGCTATCATGACGCGCGCTGACGCTGGAGCTTAGCACCTTTGGGAACGAGACTG 1194
```

```
Qy 107 ArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSer--- 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1195 ATGGACCTCTCGTGGTGCACAGCAGGATTCCCGGATGCCCATCGATGATATGTCGCCG 1254

Qy 126 -----ValLysTyrLeuArgLysIleValLysValThrAspLysValGluGlyGln 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1255 ATGGCGGTCTCGCATGGATGAAGAGCTGCTGTACTATGACGACAGCAGCGCAACGCG 1314

Qy 143 LeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1315 CTGATACCCCGCAGGACGACTGGAA-----AAGAGGTCTCAACAGGTAAAGCAACGAC 1368

Qy 163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluPro----- 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1369 GCCGTAATCAAGGACAAAAAGTTCGCGGTGGTCTC---GTAGTCGAGCCTGAAGAGGGC 1425

Qy 180 -----PheTyrSerAsnAsnGluTyrPheLeuIleVal 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1426 ATACACTTTGATGTACAGTTATGGATTTTGCAGCGCTGTATCTAGCATATAAAGGTG 1485

Qy 191 TyrAspVal---PheLysAspValArgMetValPheAlaProProSerSerValGlyLys 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1486 CGAAACCTCTCGTAGCAGACGCTCAGGTGCTTCAT-----CCGGAATGCAGAAAGAAC 1539

Qy 210 PheGlyGlyAspThrAspAsnThrMetTrpProArgHisThrGlyAspPheSerValPhe 229
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1540 ACCATCCCGCATCAACACCTGGTATGCAGAAAAACAACGGGCTTACATCGATGATA 1599

Qy 230 ArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyr 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1600 -----ATAGAGTCGCT-CCGCGACCTCGCGCTCAACTATTACAGAGCCCTCTCAA 1649

Qy 250 LysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMet 269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1650 GAGCCAGTCTATACGGA-----GGAGCAGCGCGCAGCAGTAGTACTGTGAT 1694

Qy 270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1695 CAGCAGGCCCTCAAGTGGTGTAAACGCAAGCTACGGGGTGTGGCGCCGAGAT--- 1751

Qy 290 IleGluAsnGluAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLys 309
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1752 -----ATCCCGCTGTACTTCT 1769

Qy 310 GluAlaMetSerAlaAspGlnAlaThrArgIleLysTyr-----AlaSerLysTyrAla 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1770 GCCTGCCCGCCGAGGCCACCAGCGCGTGGCGGCTATATCATCATGCAGACCATATCCCA 1829

Qy 328 GlnSerAlaAsnTyrTrpLysAsnSerIle-GlyMetAsnArgGlyLeuAlaArgLeuAs 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1830 CTGGCAGCAGATGGCGCTAAAGGTGCTGTACGGG-----GACACCGA 1871

Qy 347 pValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGln 367
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1872 TTCGCTGTTCATAAAGAAATCAGAGAGCGGCGACATCATATGATGTCGAGCAGCGCAA 1931

Qy 367 yLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1932 AAAGGAG-----CACGGC---GTCGAGCTCGAGGTGGCAAGAGTACAGGTATGTCGT 1982

Qy 397 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVal 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1983 GCTATCTAACAGGAAGAAAACTATTTCCGGGTGACAAAGTCCCGCAAGGTCGAGCTC-- 2040

Qy 407 lArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLe 427
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2040 ----- 2040

Qy 427 uLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuPr 447
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2041 -AAGGCGCTGACGGGGAAA---AAGTCGCACACGCCCGCCCTTCATAAAGAGGAGCTGTCTA 2096
```

```
Qy 447 oAlaMetLeuAspIleValArgArgIleProAlaAspLysLeuProAspIlePheLy 467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2097 TTCGCTGCTGCACATA-----CTGCTGGCTGTACAGACCCGAGGACGAGTTTGA 2144

Qy 467 sAsnVal-----IleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPh 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2145 ATCGGCAAAAGCTAAGATCTCAAGGCCATACGGGCATCCGGGAAGAGG----- 2193

Qy 484 eValPheAspLysSerValValProTyrSerAsp---LysPheHisAlaMetLeuLysSe 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2194 ---CTGAGGAGAGGGGGTCCCGCTCGCGGATCTGCGCTTCAATGTGATGAAGCAA 2249

Qy 503 rMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLy 523
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2250 GGCGCCCTCTCAA---TACGTAAAGACCGCTCCCGCAGCAC----- 2286

Qy 523 sSerValIleAlaAlaArgAlaIleGluAlaAspAlaMetAlaAsnAlaTyrAlaIl 543
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2287 -----ATACGGCGCGCAGACTGCTCGAG-----AACGCAAGGGAGGT 2324

Qy 543 eGluLysGly 546
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2325 CAAAAAGGC 2334
```

## RESULT 14

```
US-10-034-623-61
; Sequence 61, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 61
; LENGTH: 2538
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2538)
US-10-034-623-61
```

## Alignment Scores:

Pred. No.:	0.0555	Length:	2538
Score:	110.00	Matches:	121
Percent Similarity:	37.06%	Conservative:	88
Best Local Similarity:	21.45%	Mismatches:	233
Query Match:	2.96%	Indels:	123
DB:	9	Gaps:	26

US-10-008-355-2 (1-712) x US-10-034-623-61 (1-2538)

```
Qy 20 GlyValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlu-----LeuAsn 35
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 901 GCGGTGGCGGATTCGACATACCCCTGTACATGATGATGCGGGATTCGCCACGCTCCGCGAC 960

Qy 36 GlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyr 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 961 GCGGTCCATCTGGACCTGTACAGGACC-----TTCTCGAACACAGGTCTTCAGCTGTAT 1014

Qy 56 SerPhe-----AspLysProSerIleAlaAsnAlaValValIle 68
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1015 GCATTTCGGGCAAGACTATACAGATTACTCCCTGAGACAGCGTGTCCAGCGCATGCTCGGC 1074
```

Qy 69 PheGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88  
Db 1075 GAGGCAAGGTCGATTATGCGGTCTCTCGGGGATCTCACTATACACAGCTCAAA 1134  
Qy 89 HisCysGlyTyrGlyAla-----IleGlnSerGlnSerThrValAspHisAspTyrLeu 106  
Db 1135 TATTGCTATCATGACGCGCGCTGACGTGGAGCTTACACCTTTTGGGAACGAGATACTG 1194  
Qy 107 ArgAspGlyPheValSerArgThrMetGlyGluGluLeuProIleProGlyLeuSer--- 125  
Db 1195 ATGGACCTCTGGTGGTGCACAGCAGGAGTTCGCCGATGCCATCGATATGTCTCCGC 1254  
Qy 126 -----ValLysTyrLeuArgLysIleValLysValThrAspLysValIleGln 142  
Db 1255 ATGGCGCTCTGCGACTGATAGAGCCTGCTGCTACTATGACACAGCAGCGCAACGCG 1314  
Qy 143 LeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162  
Db 1315 CTGATACCCCGCAGGACGAGCTGGAA-----AAGAGGTCTCAACAGGTAAAGCAACGAC 1368  
Qy 163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluPro----- 179  
Db 1369 GCCGTAAATCAAGGCAAAAGTTCGCGGTGCTCTC-----GTAGTCGACCTGAAGAGGC 1425  
Qy 180 -----PheTyrSerAsnAsnGluTyrPheLeuIleVal 190  
Db 1426 ATACACTTTGATGTATACGTTATGAGTTTTCAGACCTGTATCTAGCATATAAAGGTG 1485  
Qy 191 TyrAspVal---PheLysAspValArgMetValPheAlaProProSerSerValGlyLys 209  
Db 1486 CGAAACCTCTCGTACGAGACCTCAGGTGCGTTCAT-----CCGGAATGCAGAAAGAAC 1539  
Qy 210 PheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe 229  
Db 1540 ACCATCCCGATACCAACCTGGTATGCACGAAAAACACAGGCTTACATCGATGATA 1599  
Qy 230 ArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyr 249  
Db 1600 -----ATAGGATCGCT-CCGCGACCTCGCGTCACTATTACAGAGCCTCTCAA 1649  
Qy 250 LysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspTyrAlaMet 269  
Db 1650 GAGCGAGTCTATACGGA-----GGAGCAGCGGCGAGCAGTAGTACTGTGAT 1694  
Qy 270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg 289  
Db 1695 CAGCCAGCCCTCAAGTGGTGTAAACCGAAGCTACGGGTGATGGCGCGGAGAT--- 1751  
Qy 290 IleGluAsnGluAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLys 309  
Db 1752 -----ATTCCGCTGTACTTTCT 1769  
Qy 310 GluAlaMetSerAlaAspGlnAlaThrArgIleLysTyr-----AlaSerLysTyrAla 327  
Db 1770 GCCTGCCCGCAGGCCACCGCGGTGCGCGGTATATCATCATGTCAGACCATATCGCA 1829  
Qy 328 GlnSerAlaAsnTyrTrpLysAsnSerIle-GlyMetAsnArgGlyLeuAlaArgLeuAs 347  
Db 1830 CTGCGACGATGGCGGTAAAGTGTCTACGGG-----GACACCCA 1871  
Qy 347 pValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGln 367  
Db 1872 TTCGTGTTTCAAGAATAACAGAGGAGCGGCGAGATCATATAGTCGAGCAGCGCAA 1931  
Qy 367 YLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387  
Db 1932 AAAGAG-----CAGGCG-----GTGAGCTCGAGGTGGACAAAGAGTACAGGTATGTGCT 1982  
Qy 387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVa 407  
Db 1983 GCTATCAACAGGAAGAAACTATTTCGGGGTGCACAAAGTCCGCGCAAGGTCGAGCTC-- 2040  
Qy 407 LArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLe 427

Db 2040 ----- 2040  
Qy 427 uLysSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuPr 447  
Db 2041 -AAGGCGCTGACGGGAAA---AAGTCGCACACGCGCCCGTTTCATAAAGAGCTGTCTTA 2096  
Qy 447 oAlaMetLeuAspIleValArgArgIleProAlaAspLysLeuProAspIlePheLy 467  
Db 2097 TTCGCTGCTCACATA-----CTGTCGCTGTACAGACCGGAGGAGTGTGA 2144  
Qy 467 sAsnVal-----IleAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPh 484  
Db 2145 ATCGGCAAAAGCTAAAGATCTCAAAAGCCATAGCGGCATCCGGGAAGAGG----- 2193  
Qy 484 eValPheAspLysSerValValProTyrSerAsp---LysPheHisAlaMetLeuLysSe 503  
Db 2194 ---CTGAGGAGAGGGGGTCCCGCTGCGGATCTGCGGTTCATGTGATGATGAAGCAA 2249  
Qy 503 rMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSerLy 523  
Db 2250 GCGGCCCTCTCAA---TACGTAAAGACCGTCCCGCAGCAC----- 2286  
Qy 523 sSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIl 543  
Db 2287 -----ATACGGCGCGGCAGACTGCTCGAG-----AACGCAAGGGAGGT 2324  
Qy 543 eGluLysGly 546  
Db 2325 CAAAAAAGGC 2334  
RESULT 15  
US-10-027-801-61  
; Sequence 61, Application US/10027801  
; Publication No. US20030054364A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DORP.002A  
; CURRENT APPLICATION NUMBER: US/10/027.801  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408.020  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 61  
; LENGTH: 2538  
; TYPE: DNA  
; ORGANISM: Cenarchaeum symbiosum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(2538)  
US-10-027-801-61  
Alignment Scores:  
Pred. No.: 0.0555 Length: 2538  
Score: 110.00 Matches: 121  
Percent Similarity: 37.06% Conservative: 88  
Best Local Similarity: 21.45% Mismatches: 233  
Query Match: 2.96% Indels: 123  
DB: 9 Gaps: 26

US-10-008-355-2 (1-712) x US-10-027-801-61 (1-2538)  
Qy 20 GlyValAlaLysAlaAspLysGlyMetTrpLeuLeuAsnGlu-----LeuAsn 35  
Db 901 GCGGTGGCGGATTCGACATACCCCTGTACATGATGCGGATTCGGCCACGCTCCGCGAC 960  
Qy 36 GlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeuTyr 55  
Db 961 GCGTCCATCTGGACCTGTACAGGACC-----TTTCGAACAGGTCTTCAGCTGTAT 1014

```
Qy 56 SerPhe-----AspLysProSerIleAlaAsnAlaValIle 68
   :::|||
Db 1015 GCATTTGGCGCAAGTATACAGATTACTCCTCGAACAGCGTGTCCAAAGCGGATCGTCGGC 1074
   :::|||
Qy 69 PheGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
   |||
Db 1075 GAGGCAAGTCGATATAGCGTGTCTCGGGGATCTCACTTATACCAAGACTGCAAAAC 1134
   |||
Qy 89 HisCysGlyTyrGlyAla-----IleGlnSerGlnSerThrValAspHisAspTyrLeu 106
   :::|||
Db 1135 TATTGCTATCATGACGCGCGCTGACCTGGAGCTTACGACCTTTGGAGAGGAGATCTG 1194
   |||
Qy 107 ArgAspGlyPheValSerArgThrMetGlyGluLeuProIleProGlyLeuSer--- 125
   |||
Db 1195 ATGGACCTCTCGTGGTGGTACAGCAGGATGCGCGGATGCCATCGATATATGTCGCCG 1254
   |||
Qy 126 -----ValLysTyrLeuArgGlyIleValLysValThrAspLysValGluGlyGln 142
   :::|||
Db 1255 ATGGCGCTCGCAGTGGATGAAGAGCGCTGCTGTACTATGAGCAGACGCGGCAACGCG 1314
   |||
Qy 143 LeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
   |||
Db 1315 CTGATACCCCGCAGGAGCTGGA-----AAGAGGTCTCAACAGGTAAAGCAACGAC 1368
   |||
Qy 163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluPro----- 179
   :::|||
Db 1369 GCCGTAATCAAGGACAAAAGTTCGCGGTGCTCTC---GTAGTCGAGCCTGAAGAGGGC 1425
   |||
Qy 180 -----PheTyrSerAsnAsnGlnTyrPheLeuIleVal 190
   :::|||
Db 1426 ATACACTTTGATGTACAGTTATGATTTTGCAGCGCTGTATCTCAGCATATAAAGGTG 1485
   |||
Qy 191 TyrAspVal---PheLysAspValArgMetValPheAlaProProSerSerValGlyLys 209
   :::|||
Db 1486 CGAACTCTCTGACGAGCGCTCAGTGGCTCAT-----CCCGAATGCAGAAAGAAC 1539
   |||
Qy 210 PheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe 229
   |||
Db 1540 ACCATCCCGGATACCAACCACTGGGTATGCACGAAACACCGGCTTACATCGATGATA 1599
   |||
Qy 230 ArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyr 249
   :::|||
Db 1600 -----ATAGGATCGCT--CCGCGACCTCGCGCTCAACTATTACAGAGCCTCTCAAA 1649
   |||
Qy 250 LysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyrAlaMet 269
   :::|||
Db 1650 GAGCCAGTCTATACGGA-----GGACGCGCGGACGAGTACTGTGAT 1694
   |||
Qy 270 ThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArg 289
   |||
Db 1695 CAGCCAGGCCCTCAAGTGTGTCTAAACGCAAGCTACGGGTGATGGCGCGCGAGAT--- 1751
   |||
Qy 290 IleGluAsnGluAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLys 309
   :::|||
Db 1752 -----ATCCCGCTGTACTTTCT 1769
   |||
Qy 310 GluAlaMetSerAlaAspGlnAlaThrArgIleLysTyr-----AlaSerLysTyrAla 327
   |||
Db 1770 GCCTGCCCGCCGAGCCACCAGCGCGGTGGCGGTATATCATCATGACAGCATATCGCA 1829
   |||
Qy 328 GlnSerAlaAsnTyrTrpLysAsnSerIle--GlyMetAsnArgGlyLeuAlaArgLeuAs 347
   |||
Db 1830 CTGCGAGCAGATGGCGGTAAGGTGCTGTACGG-----GACACCGA 1871
   |||
Qy 347 pValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGln 367
   |||
Db 1872 TTCGCTGTTCATAAAGAAATCCAGAGAGCGGACGATCATGATATAGTCGAGCACGCCAA 1931
   |||
Qy 367 YLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
   |||
Db 1932 AAGGAG-----CACGCG---GTCGAGCTCGAGGTGGACAAAGAGTACAGGTATGTCGT 1982
   |||
```

Search completed: May 23, 2003, 13:41:19  
Job time : 259.87 secs

